

COMPARACIÓ DE DIVERSOS MODELS DE SUPERVIVÈNCIA PER A FENÒMENS RECURRENTS, DAVANT LA PRESÈNCIA DE DEPENDÈNCIA D'OCURRÈNCIA I LA FALTA D'INFORMACIÓ DELS EPISODIS PREVIS

Georgina Casanovas Mateu

Tutor: Albert Navarro Giné

Universitat Autònoma de Barcelona

Treball fi de grau d'estadística

2012-2013

ÍNDEX

AGRAÏMENTS	4
1. INTRODUCCIÓ	5
1.1 ANÀLISI DE SUPERVIVÈNCIA	5
1.2 EL MODEL DE COX	6
1.3 ANÀLISI DE SUPERVIVÈNCIA DE FENÒMENS RECURRENTS	7
1.3.1 MODEL DE ANDERSEN-GILL	8
1.3.2 MODEL DE FRAGILITAT	9
1.3.3 MODEL DE PRENTICE, WILLIAM AND PETERSON	10
2. JUSTIFICACIÓ	13
3. OBJECTIU I HIPÒTESIS	13
4. METODOLOGIA	15
4.1 SIMULACIÓ	15
4.1.1 GENERACIÓ DE POBLACIONS	15
4.1.2 GENERACIÓ DE MOSTRES.....	20
4.1.3 PROGRAMARI	23
4.2 AVALUACIÓ DEL RENDIMENT DELS MODELS	23
4.2.1 MODELS CONSIDERATS	23
4.2.2 BIAIX	25
4.2.3 PRECISIÓ	25
4.2.4 COBERTURA	26
4.2.5 RISCOS PROPORCIONALS	26
5. RESULTATS	27
5.1 RENDIMENT DELS MODELS	27
6. DISCUSSIÓ.....	39
7. CONCLUSIONS	41
8. BIBLIOGRAFIA.....	43
9. ANNEX	45

AGRAÏMENTS

Al tutor del meu projecte, l'Albert Navarro, per les hores dedicades en la realització del treball, per la seva capacitat d'orientació i per tots els coneixements que m'ha transmès al llarg d'aquest temps. Per la seva paciència i per totes les estones agradables que hem compartit.

Al David Moríña, per la seva ajuda en el procés de generar les poblacions i per a qualsevol dels problemes sorgits en la part de programació d'aquest treball.

A tots els companys del departament de bioestadística de la facultat de medicina, per acollir-me i fer-me l'estància molt fàcil, i per tots els bons moments que hem compartit junts.

Finalment, agrair a tots els amics i familiars que m'han donat suport durant la realització de tot el projecte.

1. INTRODUCCIÓ

1.1 ANÀLISI DE SUPERVIVÈNCIA

S'anomena anàlisi de la supervivència ¹ al conjunt de tècniques que permeten estudiar la variable “ temps fins que ocorre un esdeveniment ” i la seva dependència d'altres possibles variables explicatives.

En ciències de la salut, els estudis longitudinals són, els que aporten informació més precisa sobre la evolució dels subjectes, però a la vegada són els més costosos i els de més complexitat en quant al seu tractament estadístic. L'anàlisi de la supervivència aplicat a les ciències de la salut està format per un conjunt de tècniques estadístiques apropiades per estudis de seguiment en els que interessa analitzar la variable temps transcorregut fins que es produeix un esdeveniment que pot presentar-se durant el període d'observació, així com l'efecte sobre la variable d'un conjunt de factors pronòstics, utilitzant per exemple, el model de regressió de riscos proporcionals de Cox que s'exposa més endavant.

El terme supervivència es deu a que en les primeres aplicacions d'aquesta tècnica l'esdeveniment terminal era el “fallo ” d'un element o la mort de l'individu. Tot i això, el temps de supervivència pot ser, per exemple, el temps transcorregut des de l'inici d'un tractament fins la desaparició del trastorn, el temps que un pacient ha estat lliure de recaigudes després d'un tractament o en general, el temps transcorregut des d'un determinat moment fins que s'ha produït l'esdeveniment d'interès.

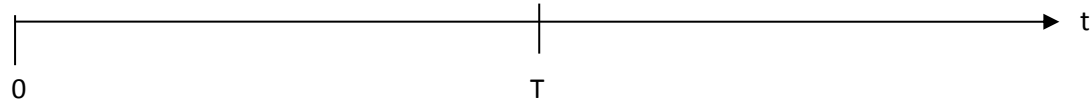
Els antecedents més llunyans de l'anàlisi de la supervivència es poden situar en la elaboració de les taules de vida que va presentar John Graunt² a Londres al 1662 i l'elaboració de les taules de mortalitat que va publicar Edmund Halley (1659- 1742)³ a partir del registre de funerals i naixements de la ciutat de Breslau. L'anàlisi de la supervivència tal i com es coneix avui en dia, té les seves arrels en la enginyeria, i està encaminat a analitzar la duració i fiabilitat dels diferents elements que formen una màquina. La segona Guerra mundial, va accelerar el desenvolupament d'aquestes tècniques per aplicar-les a la indústria militar.

En ciències de la salut, l'increment en l'ús d'aquestes tècniques començà a la dècada dels anys seixanta, especialment gràcies al mètode d'estimació de la supervivència proposat per Kaplan i Meier l'any 1958 ⁴, i en la actualitat s'ha convertit en una de les tècniques més utilitzades en medicina i epidemiologia. El motiu d'això, és, que aquesta metodologia pretén analitzar temps de supervivència valorats en unes condicions molt més flexibles que les requerides en les proves clàssiques.

L'altre impuls destacable per l'anàlisi de la supervivència es produí entre els anys 1972 i 1975 ^{5,6}, quan Sir David Roxbee Cox presentà el model de regressió més utilitzat en el context dels estudis de supervivència , el model de Cox o model de riscos proporcionals, i pocs anys després, relacionat amb aquest primer, presenta el desenvolupament de la funció de versemblança parcial.

1.2 EL MODEL DE COX

Sigui T una variable aleatòria que indica el temps fins a un esdeveniment ε (i.e, mort, recuperació d'una malaltia, diagnòstic de SIDA, etc), i sigui $x = (x_1, x_2, \dots, x_p)$ covariables fixes, és a dir, variables explicatives recollides en l'origen de l'estudi. Al conjunt de tots els factors pronòstics o covariables associades a un individu el denotarem per $x = (x_1, x_2, \dots, x_p)$ i ens referirem a ell com a perfil de l'individu.



x_1, x_2, \dots, x_p

Definim la funció de risc basal $h_0(t)$, com la funció de risc que correspondria a un individu amb valors a les covariables iguals a 0, és a dir, amb perfil $x=0$. El model general de Cox relaciona la funció de risc $h(t|X)$ en el moment t d'un individu amb perfil X amb la funció de risc el mateix moment d'un individu amb perfil $X=0$ mitjançant l'expressió:

$$h(t|X) = \psi(X)h_0(t)$$

on la funció $\psi(X)$ és no negativa i $h_0(t)$ no s'especifica paramètricament. $\psi(X)$ pot interpretar-se com el hazard ratio en el moment t d'un individu amb perfil x_i respecte un individu amb $X=0$. Es poden considerar diverses parametritzacions ψ , les més habituals en la forma log lineal, $\psi(X, B) = e^{B'X}$, la forma lineal $\psi(X, B) = 1 + B'X$ i la forma logística $\psi(X, B) = \ln(1 + e^{B'X})$. La parametrització log-lineal, és per bones raons, la més popular. Notem, que en particular, aquesta parametrització assegura que $\lambda(t|Z)$ és no negativa sense necessitat d'afegir restriccions als paràmetres.

El model bàsic de Cox estableix la següent regressió multivariada entre el risc instantani de presentar l'esdeveniment i les covariables.

$$h(t|X) = e^{B'X}h_0(t) = e^{\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p}h_0(t)$$

En aquest model se suposa que la raó entre les funcions de risc es mantenen constants al llarg del temps ja que es verifica:

$$\frac{h(t|X)}{h_0(t)} = e^{\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p}$$

On el terme de la dreta només depèn dels valors de les covariables i no del temps t . Segons aquest model, la distribució de l'error no està especificada. El factor $e^{\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p}$ correspon al risc instantani relatiu (d'ara endavant hazard ratio en la seva formulació en anglès) d'un individu amb perfil X_i respecte d'un individu amb perfil $X_i=0$ i expressa quantes vegades és més gran el risc de tenir un esdeveniment amb un perfil X que amb un perfil $X=0$.

Una forma equivalent de plantejar el model és mitjançant la transformació logarítmica. En aquest cas, el model s'expressa mitjançant l'equació:

$$\ln \left(\frac{h(t|X)}{h_0(t)} \right) = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p ,$$

I a la quantitat $\eta = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$ se la coneix com l'índex pronòstic o risk score. En aquest model no hi ha terme constant ja que aquest l'absorbeix $h_0(t)$.

Ara bé el fet que els individus puguin experimentar diversos episodis del mateix esdeveniment durant un període d'observació fa que sigui necessària l'aplicació d'una metodologia diferent de la utilitzada en l'anàlisi de supervivència estàndard per a fenòmens en què hi ha un únic episodi possible.

1.3 ANÀLISI DE SUPERVIVÈNCIA DE FENÒMENS RECURRENTS

A diferència de les tècniques clàssiques de supervivència que suposen que un esdeveniment d'interès només es pot produir una sola vegada en una mateixa observació, en l'àmbit de les ciències de la salut, existeixen múltiples fenòmens recurrents, aquells que poden repetir-se en més d'una ocasió en un mateix individu. Exemples d'aquest tipus poden ser la incapacitat laboral en treballadors o la ocurrència d'atacs d'asma.

A l'analitzar fenòmens recurrents el principal problema del model de Cox es deu a que aquest model està pensat per fenòmens que només poden passar una vegada, de manera que si les diverses ocurrencies d'un mateix episodi es tracten com independents, pot donar lloc a l'incompliment de la hipòtesi de riscos proporcionals. Aquest fet, es deu fonamentalment a l'existència simultània o per separat del següents fenòmens:

Heterogeneïtat individual: alguns individus tenen un risc més alt o inferior en presentar una ocurrencia a causa d'efectes desconeguts o no registrats.

Dependència d'ocurrencia: l'ocurrencia d'un esdeveniment modifica el risc d'una nova ocurrencia en l'individu que prèviament ja n'ha experimentat.

Els supòsits en què es basa el model de Cox, es relacionen estretament amb els de la distribució de Poisson, així, a priori, ha d'existir una estabilitat en el fenomen que s'estudia. En el cas de la Poisson, temporal i entre les observacions: el risc de tenir un episodi ha de ser constant sobre el temps i entre individus un cop fixades les seves característiques (o sigui les covariables). El model de Cox estàndard assumeix aquest últim, mentre que incorpora el primer en escala temporal i en la proporcionalitat de riscos. Un segon supòsit que s'ha de complir en la Poisson i en el model de Cox és la independència de ocurrencia: el risc d'ocurrencia d'un esdeveniment és independent al número previ d'ocurrencies de cada individu.

L'existència d'heterogeneïtat individual, situació que rep el nom de fragilitat o propensió, i/o existència de dependència d'ocurrencia, situació de contagi, implica la violació del supòsit d'independència entre els temps d'ocurrencia que estableix el model de Cox.

Amb el propòsit de solucionar aquest problema aparegueren el model de Andersen-Gill (AG)⁷⁷, que no és més que una generalització del model de Cox, però amb la diferència que la hipòtesi d'aquest model suposa que les múltiples observacions d'un mateix individu són independents, condicionades a les variables explicatives. A la pràctica, aquesta limitació pot conduir a estimadors esbiaixats i ineficients, ja que per lògica, les observacions d'un mateix individu solen estar positivament correlacionades⁸.

Amb el fi de trobar una solució apareixen a final dels anys vuitanta els models marginals i els condicionals en l'anàlisi de la supervivència multivariant.

1.3.1 MODEL DE ANDERSEN-GILL

Dins dels models de regressió condicionals, el més popular és un model basat en processos comptadors proposat per Andersen i Gill (AG)⁷.

AG adopta l'assumpció clàssica del model de riscos proporcionals de que el risc de tenir una ocurrència en un individu és independent a les ocurrències prèvies que aquest mateix individu ja ha patit. És a dir, que la funció de risc basal és comú per totes les ocurrències i les múltiples observacions de cada subjecte són independents, condicionalment al vector de covariables.

D'una altra manera, AG, considera a cada individu com un procés comptador amb processos múltiples i "increments" (temps entre esdeveniments successius), essencialment independents donada la història de totes les variables observables fins als temps d'ocurrència de l'esdeveniment,

En aquest model, la funció de risc per a l'*i*-èssim individu en el temps *t* té la següent expressió:

$$h_i(t) = h_0(t)e^{X_i(t)\beta}$$

On $h_0(t)$ és la funció de risc basal, β és un vector de paràmetres desconeguts i $X_i(t)$ és una matriu de covariables, el valor de les quals pot canviar al llarg del temps.

Aquest model és una extensió del model de Cox i posseeix dues característiques essencials:

- Tota la influència dels esdeveniments previs sobre les recurrències futures, si hi són, són el resultat de les covariables dependents del temps.
- Les covariables tenen un efecte multiplicatiu en la taxa instantània del procés comptador.

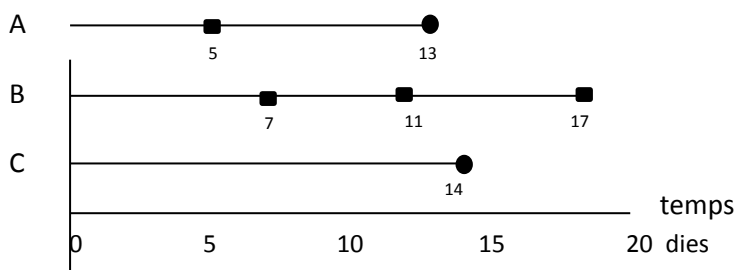
La primera de les característiques, fa que aquest model sigui també conegut com un dels models de reparacions perfectes, ja que les recurrències no depenen d'ocurrències prèvies.

La segona característica, fa que aquest model també sigui conegut com el model d'intensitat multiplicativa.

En aquest model, els individus es divideixen en diferents registres segons els intervals d'observació i cada registre es tracta com si fos un individu diferent. Les variables explicatives són fixes dins

de cada episodi però, si poden canviar el seu valor al llarg del període (per exemple, passar de ser fumador a no ser-ho, etc) , poden variar entre els registres d'un mateix individu; així, el risc que es compleixi la hipòtesi de riscos proporcionals és el risc de cada esdeveniment o recurrència per un individu determinat i aquest risc no està afectat per esdeveniments o recurrències anteriors del mateix individu.

Figura 1.Tractament dels individus segons el model d'AG.



En aquest exemple (figura 1) es pot observar tres individus diferents, l'individu A, amb un esdeveniment als 5 dies i un final d'estudi als 13 dies, l'individu B, amb 3 esdeveniments , i l'últim, individu C, on s'observa que durant els 14 dies de seguiment no ha experimentat cap episodi.

El model AG, compara doncs, aquells individus que estan sota un mateix conjunt a risc. Per exemple, als 12 dies, compararà els 3 individus, ja que aquests 3 estan a risc de patir un esdeveniment en aquest temps determinat.

1.3.2 MODEL DE FRAGILITAT

La idea de fragilitat és una forma còmode d'introduir efectes aleatoris, l' associació i l'heterogeneïtat no observada en els models de dades de supervivència. En la seva forma més simple, una fragilitat és un factor de proporcionalitat aleatòria no observada que modifica la funció de risc d'un individu o d'un conjunt d'individus. El concepte de fragilitat es remunta al treball de Greenwood y Yule (1920)⁹ sobre la " propensió a accidents". El propi terme de fragilitat va ser introduït per Vaupel (1979)¹⁰ en els models de supervivència univariant i el model va ser promogut la seva aplicació a les dades de supervivència multivariant en un article de Clayton (1978) ¹¹sobre la incidència de les malalties cròniques en famílies.

Els models de fragilitat són extensions del model de riscos proporcionals, més conegut com el model de Cox. Normalment, en la majoria d'aplicacions clíniques, l'anàlisi de supervivència suposa implícitament una població homogènia . Això significa que tots els individus de la mostra, estan subjectes, en principi, a un mateix risc si comparteixen característiques (és a dir, tenen els mateixos valors a les covariables). Moltes vegades però, no es pot suposar que aquest risc sigui homogeni per a tots els individus, sinó que hi ha diferents grups amb riscos diferents inclús compartint un conjunt de característiques.

En els models de fragilitat univariants es pretén modelitzar la dependència entre els diferents temps de recurrència introduint un únic efecte aleatori multiplicatiu, comú a totes les observacions d'un mateix individu (i , per tant, invariant en el temps), o diferent per a cadascun dels individus. Així, la fragilitat es pot entendre com una variable aleatòria no observable, i per tant, no inclosa en el model, la qual es la causant de la dependència entre les recurrències.

S'han estudiat moltes alternatives per a la distribució de probabilitat de la variable de fragilitat, com per exemple, la distribució gamma, la distribució positiva estable i la distribució gaussiana inversa, entre d'altres. Però, el més utilitzat és un model de fragilitat amb distribució de probabilitat gamma de mitjana unitària i variància desconeguda.

Existeixen diverses formulacions amb models de fragilitat, entre els que destaquem el shared frailty (fragilitat compartida) , el joint frailty (fragilitat conjunta) i unshared frailty (fragilitat individual), que és el que teòricament es podria adaptar millor per al nostre problema.

En el model de fragilitat individual, suposem que tenim una mostra de I individus on algunes observacions tenen major risc a causa de l'heterogeneïtat no observada , és a dir, variables importants per al fenomen que s'estudia, però que no es tenen en compte (normalment, perquè no es registren).

La seva funció de risc es pot especificar de la següent manera:

$$h_i(t) = h_0 v_i e^{x_i(t)\beta}$$

On v_i és un efecte aleatori individual que normalment s'especifica mitjançant una distribució gamma amb $E(v_i) = 1$ i $\text{Var}(v_i) = \theta$. Donat que el model de fragilitat capta variabilitat causada per variables desconegudes, en el cas de fenòmens recurrents amb presència de dependència d'ocurrència i desconeixement dels episodis previs soferts per cada individu, es podria hipotitzar que part de la variabilitat introduïda pel fet de desconèixer els episodis previs de cada individu es tindria en compte mitjançant el terme de fragilitat.

1.3.3 MODEL DE PRENTICE, WILLIAM AND PETERSON (PWP)

El model de Prentice, Williams i Peterson, mètode (PWP, 1981)¹² és un model marginal respecte l'estimació de paràmetres, però condicional respecte la construcció del conjunt d'individus a risc. Aquest model permet que el risc basal variï entre diferents ocurrències. Es tracta, doncs, d'un model de riscos proporcionals amb estrats dependents del temps, on la dependència entre recurrències es controla estratificant pel nombre de vegades que l'esdeveniment d'interès ha passat anteriorment.

Així doncs, el model PWP inclou en el conjunt de risc per la recurrència k -èssima, únicament els individus que han experimentat $k-1$ recurrències.

Prentice, William i Peterson proposen dos models semi-paramètrics de riscos proporcionals:

Temps des de l'inici de l'estudi, on la funció de risc subjacent s'inclou com una funció del temps des de l'inici de l'estudi:

$$h_{ik}(t) = h_{0k}(t)e^{X_i(t)\beta}$$

Temps des de la última recurrència, on la funció de risc subjacent és una funció del temps des de la recurrència anterior al temps de la fallada.

$$h_{ik}(t) = h_{0k}(t - t_{k-1})e^{X_i(t)\beta}$$

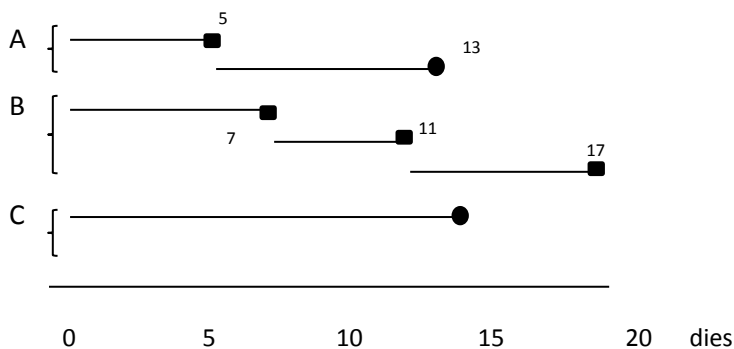
S'observa que, en definitiva, es tracta de models AG estratificats, degut a que tenen un conjunt de risc condicional i el risc basal és específic per cada recurrència. Aquest model es diferencia de AG en dos aspectes:

El conjunt de risc per a la k – èssima recurrència es limita a individus que han experimentat les k-1 primeres recurrències. Així doncs, la funció de risc subjacent i els paràmetres de la regressió poden variar entre les diferents recurrències.

El principal problema de PWP es troba en la forma de construir els conjunts a risc, ja que comporta una pèrdua d'aleatorietat, donat que els individus amb més risc van sortint de l'estudi. Aquesta pèrdua serà més gran a mesura que ens anem movent a estrats superiors. En aquests casos, la consistència de les estimacions depèn de la inclusió de totes les variables rellevants.

Evidentment, a priori, aquests models haurien de ser els més escaients en el cas de presència de dependència d'ocurrència. A l'estratificar per l'episodi a risc (k) s'obté un risc basal específic que pretén acomodar la dependència d'ocurrència. El problema però, que justifica el treball, és quan el valor de k és desconegut, situació freqüent en algunes disciplines com la salut pública.

Figura 2. Tractament dels individus segons el model de PWP



Si observem ara, l'exemple de com tracta els individus el model PWP respecte el model d'AG veiem que és força diferent. Si mirem la figura 2, veiem la mateixa representació que en el model AG, però estratificat. En aquest cas, el model PWP, al dia 9 per exemple, no compararà els 3 individus, sinó que compararà només l'individu A amb l'individu B, ja que aquests són els que estan a risc de patir un segon episodi, en canvi, no es podrà comparar amb l'individu C, ja que aquest, encara no ha presentat el primer episodi.

2. JUSTIFICACIÓ

En l'àmbit mèdic o de les ciències de la salut, sol ser freqüent l'interès per l'anàlisi de la supervivència en fenòmens recurrents. En aquest cas, és necessari utilitzar una metodologia diferent a la utilitzada en l'anàlisi de supervivència estàndard. El principal problema que comporta els estudis d'aquest tipus de fenòmens és que els diversos episodis d'un mateix individu estan correlacionats i el risc basal pot ser diferent en funció dels episodis previs que s'hagin patit.

El model PWP sembla ser el model més adequat en aquest context: té en compte que els episodis d'un mateix individu no són independents, per una banda, i permet estimar un risc basal específic segons el nombre d'episodis previs que s'han presentat per altra. Però sovint, les ocurrències prèvies dels esdeveniments d'interès que ha tingut un individu no són conegudes i per tant, el model PWP no es pot aplicar.

El que es pretén en aquest treball és, doncs, valorar el rendiment de les diverses alternatives al model PWP en el context de l'anàlisi de supervivència per a fenòmens recurrents, quan es desconeix el nombre d'ocurrències prèvies de l'esdeveniment d'interès.

3. OBJECTIU I HIPÒTESIS

OBJECTIU:

Valorar el rendiment de les diverses alternatives al model PWP en el context de l'anàlisi de la supervivència per a fenòmens recurrents, quan es desconeix el nombre d'ocurrències prèvies de l'esdeveniment d'interès.

HIPÒTESIS:

- 1- Els models amb risc basal comú sobreestimen els coeficients en una situació de dependència d'ocurrència creixent.
- 2- Els models de fragilitat podrien captar part de l'efecte de la dependència d'ocurrència a través de l'efecte aleatori individual.
- 3- Els models basats en el temps d'exposició compleixen amb major freqüència la proporcionalitat de riscos en relació als models anàlegs basats en el temps d'estudi.
- 4- La consideració del temps d'exposició com a covariable podria permetre una estimació menys esbiaixada dels coeficients.
- 5- A major dependència d'ocurrència pitjor rendiment dels models examinats en aquest treball
- 6- Quan major sigui el valor del paràmetre beta, major la sobreestimació de l'efecte per part dels models amb risc basal comú.

4. METODOLOGIA

4.1 SIMULACIÓ

4.1.1 GENERACIÓ DE POBLACIONS

Mitjançant el paquet *complex.surv.data.sim* de l'R ¹³, s'han creat 3 escenaris diferents, cadascun amb un total de 250.000 individus i un seguiment de 20 anys (7305 dies). El que determina la diferència entre les poblacions generals és la $h_{0k}(t)$, la variació de la qual permet construir poblacions on el hazard basal, i per tant, el hazard ratio en funció dels episodis previs, és distint. Això fa que existeixi una dependència d'ocurrència, on l'ocurrència d'un esdeveniment modifica el risc d'una nova ocurrència en l'individu que prèviament ja n'ha experimentat.

Cadascuna de les funcions s'han creat a partir de 3 exponencials, on a partir d'aquí hem pogut calcular el hazard per a cadascun dels episodis.

El hazard s'expressa com:

$$h_{ik}(t) = \exp(\beta_{0k} + \beta_1 X_1 + \beta_2 X_2) x Z_i$$

On $h_{0k}(t) = \exp(-\beta_0)$ i Z_i és una covariable contínua que segueix una distribució Normal amb mitjana 1 i desviació típica 0.2, que s'utilitza per introduir heterogeneïtat individual.

En totes les simulacions realitzades s'ha fixat que $\beta_1 = 0.2$, $\beta_2 = 0.7$, essent aquests paràmetres sempre iguals en les poblacions generades i independents de l'episodi al que s'està a risc. Assumim també en cadascuna de les poblacions simulades, que el 100 % dels individus no estava en risc abans del seguiment.

Un cop obtingudes les poblacions, hem fet una petita descriptiva del episodis i la densitat d'incidència, per cadascuna de les poblacions.

La incidència és el nombre de casos nous en una determinada malaltia o esdeveniment que apareixen en un període de temps i expressa la força en què una malaltia té per canviar l'estat de salut d'una població a l'estat d'una malaltia per unitat de temps en relació a la població susceptible en aquest moment.

Noti que en les poblacions generades s'ha establert una dependència d'ocurrència creixent, és a dir, s'ha suposat que el risc d'experimentar l'esdeveniment augmenta a mesura que es van presentant episodis. S'ha escollit aquest patró donat que és el que habitualment es troba en els fenòmens relacionats amb la salut.

Població 1

beta	Hazard (taxa/dia)	Hazard (taxa/setmana)	HR
$\beta_{01} = 8.109$	0.0003	0.0021	
$\beta_{02} = 7.927$	0.0004	0.0025	1.20
$\beta_{03} = 7.521$	0.0005	0.0038	1.80

La primera població, es caracteritza per tenir la dependència d'ocurrència més baixa entre les poblacions generades. Veiem que el risc de patir el primer episodi és de 0.0003 per individu/ unitat de temps (que suposa que és un dia), i el segon, un cop l'individu ja n'ha experimentat un és de 0.0004. Així doncs, el hazard ratio de patir el segon episodi envers el primer és de 1.20 (o dit d'altra manera, el risc basal de patir l'esdeveniment és un 20% major). Si mirem ara, el risc de patir un tercer episodi quan s'ha observat el segon, creix fins a 0.0038, amb un hazard ratio de 1.80 en relació al primer episodi.

Taula 1: casos observats població 1

Nº episodis totals	N	% casos
0	99336	39.73
1	62533	25.01
2	34882	13.95
3	20422	8.17
4	12464	4.99
5	7694	3.08
6	4785	1.91
7	2986	1.19
8	1754	0.70
9	1205	0.48
>=10	1939	0.78

Es pot observar a la (taula 2) població 1 que el 39.73 % dels individus no pateixen cap episodi durant el seguiment, i un 25.01 % només 1, de manera que més de la meitat de la població com a molt, haurà tingut un episodi.

Taula 2: densitat incidència població 1

Nº episodis a risc	Densitat incidència	IC (95%)
1	3.68	(3,66 – 3.70)
2	4.95	(4,92 – 4.99)
3	8.42	(8,35 – 8.49)
4	9.22	(9,12 – 9.32)
5	9.73	(9,60 – 9.86)
6	10.19	(10,01 – 10.37)
7	10.88	(10,64 – 11.12)
8	11.34	(11,02 – 11.66)
9	11.92	(11,51 – 12.34)
>=10	12.97	(12,62 – 13.32)

* densitat incidència per 10.000 persones/dia

Si ens fixem ara, en la densitat d'incidència (taula 2), s'observa que aquesta és creixent si el nombre d'episodis en risc també creix. Es pot observar que la densitat d'incidència creix fortament durant els 3 primers episodis a risc, i a partir del quart, aquest comença a mantenir-se constant. Per exemple, veiem que en un dia, per cada 10.000 individus a risc del primer episodi en un dia 3.68 tindrien un episodi. En canvi, si observem la densitat d'incidència quan el nombre d'episodis és >10 el nombre de casos ja serien 12.

Població 2

beta	Hazard (taxa/dia)	Hazard (taxa/setmana)	HR
$\beta_{01} = 8.109$	0.0003	0.0021	
$\beta_{02} = 7.416$	0.0006	0.0042	2.00
$\beta_{03} = 6.500$	0.0015	0.0105	5.00

Si comparem ara la segona població envers la primera, veiem que la dependència d'ocurrència creix considerablement, es pot observar, que el risc de patir un primer episodi és de 0.0003, igual que en la primera població, mentre que patir un segon episodi quan ja se n'ha patit un, duplica aquest risc amb 0.0006. Si observem finalment, el risc de patir un tercer episodi el risc ja és 5 vegades superior al risc de presentar el primer.

Veiem doncs, que en aquest cas, aquesta dependència, és molt més elevada que en la primera població.

Taula 3 : casos observats població 2

Nº episodis totals	N	% casos
0	99627	39.85
1	54459	21.78
2	31457	12.58
3	20359	8.14
4	13551	5.42
5	9094	3.64
6	6188	2.48
7	4319	1.73
8	3011	1.20
9	2194	0.88
>=10	5741	2.30

Es pot observar que el 39.85 % dels individus no pateixen cap episodi durant els 20 anys de seguiment, i un 21.78 % només 1, de manera que més de la meitat de la població com a molt, haurà tingut un episodi.

Taula 4: densitat incidència població 2

Nº episodis a risc	Densitat incidència	IC (95%)
1	3,67	(3,65 – 3.69)
2	8,78	(8,73 – 8.84)
3	25,55	(25,35 – 25.75)
4	27,05	(26,80 – 27.30)
5	28,43	(28,11 – 28.75)
6	29,60	(29,20 – 30.00)
7	31,04	(30,55 – 31.53)
8	32,50	(31,90 – 33.11)
9	34,39	(33,63 – 35.15)
>=10	37,71	(37,21 – 38.20)

* densitat incidència per 10.000 persones/dia

Si ens fixem ara en la densitat d'incidència, s'observa que aquesta és creixent si el nombre d'episodis en risc també creix. Veiem doncs, que en un dia, per cada 10.000 individus, 3.67 tindrien un episodi. En canvi, si observem la densitat d'incidència quan el nombre d'episodis és >10 ens trobem un nombre bastant més elevat d'individus, amb un total de 37,71.

Població 3

beta	Hazard (taxa/dia)	Hazard (taxa/setmana)	HR
$\beta_{01} = 8.109$	0.0003	0.0021	
$\beta_{02} = 7.010$	0.0009	0.0063	3.00
$\beta_{03} = 5.912$	0.0027	0.0189	9.00

Finalment, la tercera població és la que presenta una dependència d'ocurrència més gran. En aquest cas, veiem que el risc entre un episodi i l'altre es triplica cada vegada fins a obtenir un HR de 9, entre el risc de tenir el tercer episodi i el primer.

Taula 4: casos observats població 3

Nº episodis totals	N	% casos
0	99180	39.67
1	51979	20.79
2	33079	13.23
3	21141	8.46
4	13897	5.56
5	9277	3.71
6	6258	2.50
7	4379	1.75
8	2995	1.20
9	2038	0.82
>=10	5777	2.31

Si observem ara taula de percentatge de casos segons els episodis (taula 5), observem, que gairebé un 40 % dels individus no han patit cap episodi , i un 20.79% només un durant els 20 anys de seguiment, però la quantitat de repetidors augmenta.

Taula 6: densitat incidència població 3

Nº episodis a risc	Densitat incidència	IC (95%)
1	3,68	(3,66 – 3.70)
2	13,90	(13,82 – 13.99)
3	62,98	(62,50 – 63.47)
4	66,72	(66,10 – 67.34)
5	69,88	(69,10 – 70.66)
6	73,43	(72,45 – 74.41)
7	76,58	(75,36 – 77.80)
8	78,63	(77,15 – 80.11)
9	83,03	(81,19 – 84.87)
>=10	92,65	(91,46 – 93.85)

* densitat incidència per 10.000 persones/dia

Si observem ara les densitats d'incidència en la població 3 (taula 6), s'observa que creix considerablement a partir del tercer episodi. Per exemple, si observem la incidència quan està en risc el segon episodi, s'observa que la taxa d'incidència és de 3,68 per cada 10.000 persones/dia . En canvi, per al tercer episodi, s'observa que aquesta taxa s'ha quadruplicat, ja que ara, per cada 10.000 individus, 62,98 tindrien tres episodis.

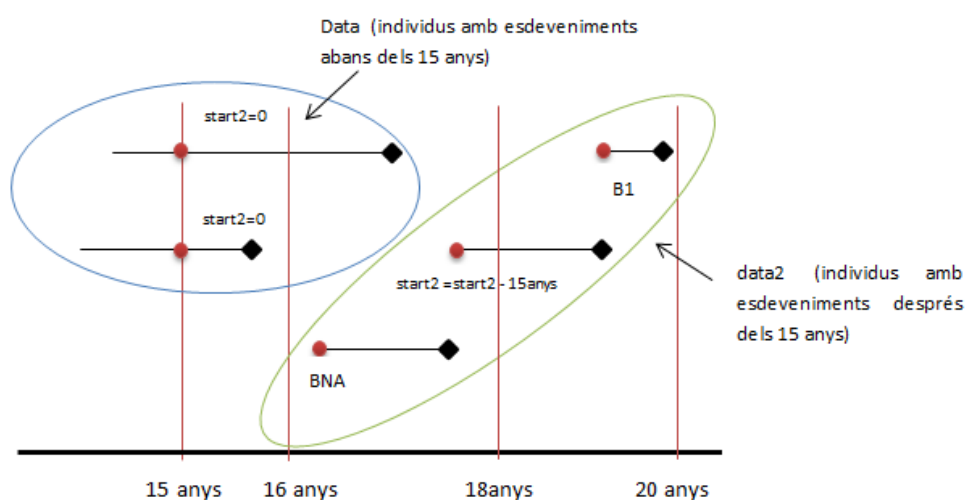
Els hazards usats en aquest treball s'han escollit de forma que puguin reproduir un esdeveniment real. Concretament, el risc basal del primer episodi de cada població, 0.0021 episodis per unitat d'observació-setmana, es correspon al hazard que troben Koopmans et al.¹⁴, per al risc de patir una incapacitat laboral de llarga durada en una cohort de 53.830 treballadors holandesos. Els hazards ratios utilitzat entre els riscos basals segons episodis previs, tot i que arbitraris, podrien correspondre precisament a episodis d'incapacitat laboral de diversos motius, tal i com descriuen Reis et al.¹⁵ en una cohort de treballadors d'un hospital universitari.

4.1.2 GENERACIÓ DE MOSTRES

D'entrada, per a cadascuna de les poblacions generades en el punt anterior, el que volíem, era agafar 3 tipus de dissenys d'estudis diferents, corresponents a diferents talls durant els 20 anys de seguiment.

El que volíem aconseguir era el seguiment de diferents individus durant un període de temps, però sense saber el nombre d'episodis previs (k) que havien obtingut, de manera que vam considerar com el nostre punt de tall 0, els 15 anys, obtenint així, individus en que havien experimentat episodis, però que nosaltres desconexíem. Un cop establert els 15 anys com el començament de seguiment, es va obtenir els altres talls, corresponents als 16, 18 i 20 anys, de manera que així, es van crear 3 bases de dades diferents per a cadascuna de les poblacions. Una amb 1 any de seguiment (inici de seguiment als 15 anys, final als 16), l'altra amb 3 anys de seguiment (inici de seguiment als 15 anys, final als 18), i finalment una base de dades amb 5 anys de seguiment (inici de seguiment als 15 anys, final als 20), amb això, s'aconsegueix tenir tres cohorts d'una mateixa població amb temps de seguiment distint, amb individus que s'incorporen durant el seguiment (a partir dels 15 anys) i altres que quan comencem a seguir-los (al nostre temps 0, als 15 anys reals d'exposició) ja tenien una història prèvia on havien pogut presentar múltiples ocurrències de l'esdeveniment d'interès, si bé a la pràctica aquests episodis eren desconeguts.

Figura 3. Procés per obtenir les poblacions



Els temps, es van haver de re-escalar segons els diferents talls, així doncs, mitjançant un funció pròpia (veure annex), vàrem re-calcular tots els temps, de manera que el temps inicial als 15 anys fos el 0, i l'stop màxim de cada base de dades, fos el temps de seguiment màxim.

Per veure més clar aquest procediment, podem veure un petit exemple a continuació:

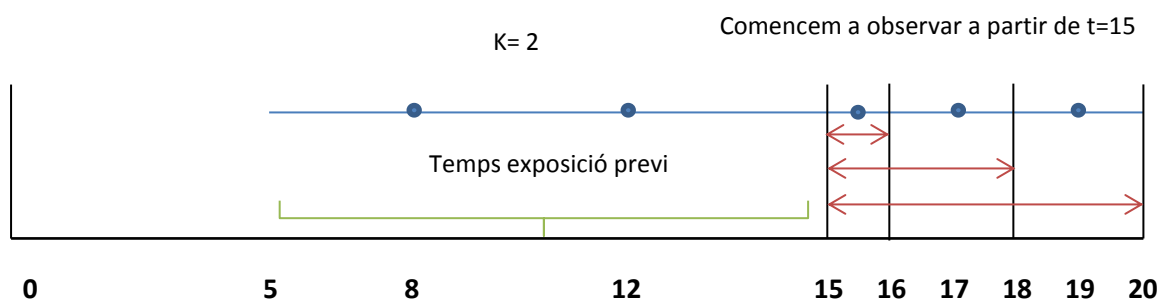
Suposem que tenim un individu que ha començat el seguiment als 5 anys i s'ha seguit fins al final de l'estudi (figura 4). El que volem, és obtenir una població on no observem alguns esdeveniments previs (k), de manera que establim l'inici de seguiment als 15 anys. Observem doncs, que aquest individu, ha tingut 2 episodis previs, que nosaltres, més endavant desconexirem .

Si agafem el tall als 16 anys, el seu primer episodi serà al $t=0.5$ ja i no al 15.5, ja que li hem de restar els 15 anys de seguiment, i acabaria als 16 anys, perquè és el màxim que observem.

En alguns dels models proposats en aquest treball, hem tingut en compte però, el temps d'exposició inicial de cada individu, en aquest cas, seria 10 anys, que són els anys que han passat entre el començament de seguiment i els 15 anys.

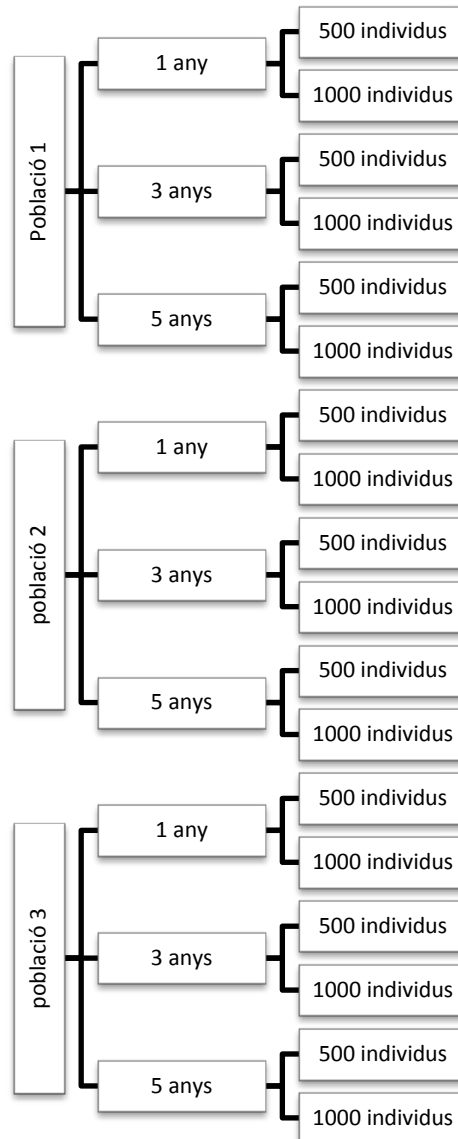
En el cas dels talls als 18 i 20 anys, el començament de l'episodi seria el mateix però el final, correspondria com a màxim al final del seguiment.

Figura 4: exemple obtenció dels diferents talls



Un cop obtingudes les subpoblacions, vam crear un programa per agafar les diferents mostres (veure annex, *Generació de mostres amb els models considerats*) , de manera que vam agafar 500 mostres de cada subpoblació amb 500 i 1000 individus. finalment, es va obtenir un resultat final de 6 bases de dades diferents per a cadascuna de les poblacions de l'estudi, 3 per cada tall (1 any, 3 anys, 5 anys) i cadascuna amb 500 i 1000 individus. (figura 5)

Figura 5 : obtenció de les diferents cohorts



Les poblacions s'han generat amb el paquet *complex.surv.data.sim* de l'R¹³. Aquest paquet, cobreix diferents situacions, incloent esdeveniments recurrents i esdeveniments múltiples. La rutina principal de la simulació permet a l'usuari introduir un nombre arbitrari de distribucions, cadascun corresponent a un nou esdeveniment o episodi, amb els seus paràmetres, escollint entre una Weibull (i una exponencial com un cas particular), log-logística i una distribució log-normal

Per a les simulacions, s'ha utilitzat un programa propi, creat especialment per aquest treball, que s'especifica en l'annex. Dins d'aquest programa, s'usa l'ajust de models de supervivència per a fenòmens recurrents que es troben inclosos en el paquet *survival* de l'R¹⁶.

4.2 AVALUACIÓ DEL RENDIMENT DELS MODELS

4.2.1 MODELS CONSIDERATS

Els models considerats en aquest treball es mostren a la taula 7. Si bé en la situació plantejada en què es desconeix el nombre d'episodis previs és impossible usar el model de PWP, en aquest cas, hem calculat com a referència i validació en les mostres simulades.

Taula 7. Models considerats

MODEL	FUNCIÓ DE RISC
PWP	$h_{ik}(t) = h_{0k}(t)e^{X_i\beta}$
PWPZ	$h_{ik}(t) = h_{0k}(t)e^{X_i\beta+z_i}$
AGc	$h_i(t) = h_0(t)e^{X_i\beta}$
AGc1	$h_i(t) = h_0(t)e^{X_i\beta+z_i}$
AGe	$h_i(t') = h_0(t')e^{X_i\beta}$
AGe1	$h_i(t') = h_0(t')e^{X_i\beta+z_i}$
FRAc	$h_i(t) = h_0(t)v_i e^{X_i\beta}$
FRAc1	$h_i(t) = h_0(t)v_i e^{X_i\beta+z_i}$
FRAe	$h_i(t') = h_0(t')v_i e^{X_i\beta}$
FRAe1	$h_i(t') = h_0(t')v_i e^{X_i\beta+z_i}$
AGexp	$h_i(t) = h_0(t)e^{X_i\beta+t'_{ini}}$
AGexp1	$h_i(t) = h_0(t)e^{X_i\beta+z_i+t'_{ini}}$
FRAexp	$h_i(t) = h_0(t)v_i e^{X_i\beta+t'_{ini}}$
FRAExp1	$h_i(t) = h_0(t)v_i e^{X_i\beta+z_i+t'_{ini}}$

On $h_0(t) = e^{-\beta_0}$, $X_i\beta$ són el vector de covariables i coeficients, z_i és la heterogeneïtat individual de cada individu i v_i la fragilitat.

Recordem que el temps d'exposició inicial correspon al temps real, que va començar l'individu a la població. Observem doncs, per exemple, la diferència entre el model AG i AGe, recau en què el model AG fixa el temps 0 en els 15 anys, mentre que el model AGe utilitza com escala temporal l'exposició, de tal manera, que l'individu representat a la figura 6 comença el seu seguiment als 5 anys i així consta en les dades, on el seu primer registre comença en aquest temps, no el zero.

La comparació dels resultats de l'ajust dels models amb els valors reals utilitzats per a simular les dades, proporciona una mesura del rendiment i la precisió associada al procés de simulació. Les mesures més utilitzades per comprovar l'acompliment són l'avaluació del biaix, precisió i cobertura. Collins¹⁷, va posar de relleu la importància d'examinar més d'un criteri d'acompliment com ara l'error quadràtic mitjà (MSE), la cobertura i l'amplitud dels intervals de confiança, a més de biaix, ja que els resultats poden variar segons els criteris. En general, l'expectativa de les estimacions simulades és d'interès principal, i per tant, la mitjana de les estimacions de totes les estimacions s'utilitza per calcular la precisió de mesures, com ara el biaix. En jutjar l'acompliment dels diferents mètodes, hi ha una compensació entre la quantitat de tendència i la variabilitat.

24

Criteri d'avaluació	Fórmula
BIAIX	
Biaix	$\delta = \bar{\beta} - \beta$
Percentatge biaix	$\delta = \left(\frac{\bar{\beta} - \beta}{\beta} \right) * 100$
PRECISIÓ	
MSE (error quadràtic mitjà)	$(\bar{\beta} - \beta)^2 + SE(\hat{\beta})^2$
COBERTURA	Proporció de vegades que 100(1- α)% interval de confiança
PROPORTIONAL HAZARD	
Ph	Proporció de vegades que es compleix PH >0.05

4.2.2 BIAIX

El biaix és la desviació en una estimació de la quantitat veritable, que pot indicar el rendiment dels mètodes que estan sent avaluats. Una avaluació del biaix, és la diferència entre l'estimació de la mitjana i el valor real, $\delta = \bar{\beta} - \beta$ (taula 1). La quantitat de biaix que es pot considerar problemàtic pot variar entre $\frac{1}{2}SE(\hat{\beta})$ i $2SE(\hat{\beta})$. Una altra opció és calcular el biaix com un percentatge del valor real (taula 8) que proporciona el valor vertader no igual a zero.

4.2.3 PRECISIÓ

L'error quadràtic mitjà (MSE) d'un estimador és una de les maneres de quantificar la diferència entre els valors implícits en un estimador i els valors veritables de la quantitat que s'estima. L'MSE és una funció de risc que correspon al valor de la pèrdua de l'error al quadrat o pèrdua quadràtica. L'MSE mesura la mitjana dels quadrats dels errors.

L'MSE proporciona una mesura útil de la precisió global, ja que incorpora les dues mesures de biaix i la variabilitat. L'arrel quadrada de l'MSE transforma la part posterior de l'MSE en la mateixa escala que el paràmetre.

La llargada mitjana de l'interval de confiança del 95% de l'estimació dels paràmetres es considera sovint com a una eina d'avaluació en els estudis de simulació. Si les estimacions dels paràmetres són relativament no esbiaixades, els intervals de confiança més estrets impliquen estimacions més precises, el que comporta millors resultats en l'eficiència i la potència estadística.

4.2.4 COBERTURA

La cobertura d'un interval de confiança és la proporció de vegades que l'interval de confiança obtingut conté el valor del paràmetre real. Un excés de cobertura, on les taxes de cobertura estan per sobre del 95%, suggereix que els resultats són massa conservadors. En canvi, quan les taxes de cobertura són inferiors al 95% , és inacceptable, ja que això indica un excés de confiança en les estimacions. Un possible criteri per a l'acceptabilitat de la cobertura, és que la cobertura no ha de caure aproximadament fora de $2SE$ de la probabilitat de cobertura nominal (p), $SE(p) = \sqrt{p(1 - p)/B}$, o n/B equival al nombre de simulacions.

4.2.5 RISCOS PROPORCIONALS

Una assumpció clau per poder utilitzar el model de Cox (en el qual es basen tots els models usats en aquest treball) és que es compleixi la propietat de riscos proporcionals. Hi ha varies formes de verificar la propietat de riscos proporcionals de les dades. En el nostre cas, per estudiar aquesta propietat, utilitzarem un test basat en residus de Schoenfeld ¹⁸.

5. RESULTATS

5.1 RENDIMENT DELS MODELS

A les següents taules, podem observar les estimacions de les β per a cadascun dels models proposats, juntament amb els criteris d'avaluació esmentats anteriorment, com el biaix, al cobertura, la longitud de l'interval, l'error quadràtic mitjà i el percentatge d'individus que compleixen l'assumpció de riscos proporcionals.

Taula 9. Resultats població 1 , 1 any, n=500

Model	Estimació β_1	Estimació β_2	% biaix β_1	% biaix β_2	Cobertura β_1	Cobertura β_2	long β_1	long β_2	MSE β_1	MSE β_2	PH β_1	PH β_2
PWP	0.191	0.650	-4.715	-7.157	94.0	92.2	0.391	0.419	0.010	0.014	95.0	94.6
PWPZ	0.198	0.667	-0.977	-4.745	93.4	92.6	0.383	0.413	0.009	0.012	95.4	95.0
AGc	0.245	0.836	22.379	19.362	92.4	82.4	0.465	0.480	0.016	0.034	88.6	87.2
AGc1	0.248	0.838	23.756	19.782	92.8	80.4	0.451	0.469	0.015	0.034	89.2	89.2
AGe	0.308	1.055	54.042	50.692	87.0	22.6	0.494	0.515	0.028	0.143	78.0	57.0
AGe1	0.313	1.063	56.367	51.789	86.2	18.0	0.474	0.498	0.028	0.148	80.2	57.0
FRAc	0.245	0.836	22.267	19.455	93.4	83.0	0.468	0.480	0.016	0.034	91.0	89.6
FRAc1	0.247	0.838	23.672	19.668	92.8	81.0	0.456	0.469	0.015	0.034	91.6	90.6
FRAe	0.309	1.058	54.386	51.190	88.0	23.8	0.508	0.519	0.029	0.147	92.8	68.8
FRAe1	0.312	1.061	56.234	51.608	77.2	12.2	0.402	0.426	0.028	0.147	93.8	69.2
AGexp	0.252	0.860	26.118	22.853	91.8	74.0	0.454	0.471	0.016	0.040	90.4	88.8
AGexp1	0.255	0.863	27.541	23.357	92.8	72.2	0.439	0.458	0.015	0.041	91.6	89.6
FRAexp	0.289	0.985	44.542	40.732	89.0	37.2	0.467	0.490	0.022	0.098	93.4	93.4
FRAExp1	0.293	0.987	46.325	40.965	87.8	33.0	0.447	0.474	0.022	0.098	93.7	93.6

Com sabem des d'un principi, els millors models que estimen millor, són els models de Prentice, William and Peterson en què té en compte els episodis previs que han tingut els individus. És per això, que volem comparar, si amb els altres models proposats podríem obtenir bons resultats per estimar aquestes betes.

En la població 1, amb 1 anys i 500 mostres de 500 individus (taula 9), es pot observar que la millor alternativa seria utilitzar el model de fragilitat. S'observa que tot i tenir un biaix del 20% la cobertura i l'assumpció de riscos proporcionals és proper al 90%. Aquest model obté resultats lleugerament millors als d'AG. Per altra banda, es pot observar que els pitjors models, serien els models d'AG amb temps d'exposició.

La cobertura mostra valors clarament diferents, en general, segons la beta estimada. Així, la beta major, aquella que representa un major efecte de la covariable, mostra taxes de cobertura clarament inferiors.

Taula 10. Resultats població 1 , 1 any, n=1000

Model	Estimació β_1	Estimació β_2	% biaix β_1	% biaix β_2	Cobertura β_1	Cobertura β_2	long β_1	long β_2	MSE β_1	MSE β_2	PH β_1	PH β_2
PWP	0.188	0.641	-6.011	-8.393	94.8	86.6	0.277	0.296	0.005	0.009	93.6	93.6
PWPZ	0.191	0.658	-4.373	-5.989	95.8	90.2	0.271	0.292	0.005	0.007	94.2	93.8
AGc	0.240	0.823	19.930	17.572	92.2	68.8	0.329	0.339	0.009	0.023	87.2	82.6
AGc1	0.239	0.826	19.381	18.052	91.8	65.6	0.319	0.331	0.008	0.023	89.8	83.4
AGe	0.303	1.039	51.373	48.467	78.4	5.0	0.350	0.364	0.019	0.124	71.6	36.4
AGe1	0.302	1.048	51.190	49.731	78.4	3.4	0.336	0.352	0.018	0.129	75.6	39.6
FRAc	0.240	0.824	20.228	17.731	92.4	67.4	0.332	0.339	0.009	0.023	89.0	83.8
FRAc1	0.238	0.826	19.194	17.797	91.6	66.0	0.322	0.331	0.008	0.023	90.2	85.0
FR Ae	0.303	1.043	51.345	49.052	80.6	5.4	0.360	0.367	0.019	0.127	89.0	46.0
FR Ae1	0.301	1.046	50.280	49.475	67.2	1.0	0.283	0.299	0.018	0.129	88.6	47.2
AGexp	0.247	0.848	23.374	21.132	92.4	56.6	0.321	0.333	0.009	0.029	88.8	85.4
AGexp1	0.246	0.852	22.929	21.658	91.6	52.6	0.310	0.323	0.008	0.030	91.0	85.2
FR Aexp	0.284	0.973	42.217	39.069	84.8	12.8	0.331	0.347	0.014	0.083	92.8	92.0
FR Aexp1	0.282	0.975	41.208	39.325	83.8	11.2	0.317	0.335	0.013	0.083	93.3	92.5

Si mirem ara, la mateixa població, amb el mateix tall, però amb un total de 1000 individus (taula 10), es pot observar pràcticament els mateixos resultats que en la taula anterior. S'observa doncs, que els models de fragilitat, seria la millor alternativa, per a poder analitzar aquest tipus de dades.

Taula 11. Resultats població 1 , 3 anys, n=500

Model	Estimació β_1	Estimació β_2	% biaix β_1	% biaix β_2	Cobertura β_1	Cobertura β_2	long β_1	long β_2	MSE β_1	MSE β_2	PH β_1	PH β_2
PWP	0.185	0.652	-7.465	-6.861	95.4	90.4	0.390	0.416	0.010	0.014	95.0	93.0
PWPZ	0.189	0.669	-5.447	-4.498	95.0	93.2	0.382	0.411	0.009	0.013	95.2	93.6
AGc	0.236	0.832	18.095	18.896	93.8	80.8	0.463	0.477	0.015	0.033	88.2	86.8
AGc1	0.236	0.835	17.805	19.240	94.4	80.2	0.450	0.467	0.013	0.033	89.4	87.2
AGe	0.295	1.052	47.468	50.287	89.0	24.0	0.493	0.513	0.025	0.142	77.2	52.4
AGe1	0.295	1.060	47.526	51.472	89.2	20.0	0.473	0.496	0.023	0.147	79.4	51.8
FRAc	0.237	0.833	18.637	19.059	94.2	79.8	0.467	0.478	0.015	0.034	91.2	87.4
FRAc1	0.236	0.835	18.149	19.261	94.4	79.8	0.455	0.467	0.014	0.033	91.4	87.8
FR Ae	0.296	1.056	47.928	50.915	89.6	25.4	0.508	0.518	0.026	0.147	90.4	66.4
FR Ae1	0.295	1.059	47.510	51.328	80.0	13.2	0.402	0.425	0.024	0.148	90.4	64.4
AGexp	0.242	0.859	21.056	22.773	93.8	72.6	0.452	0.469	0.015	0.041	88.8	86.2
AGexp1	0.242	0.863	20.782	23.214	94.8	71.0	0.438	0.456	0.013	0.041	91.2	87.2
FR Aexp	0.279	0.985	39.695	40.734	91.4	38.4	0.467	0.490	0.021	0.099	93.0	94.2
FR Aexp1	0.278	0.987	39.215	41.003	89.6	33.2	0.447	0.473	0.019	0.099	93.4	94.6

En el cas d'observar la població 1, però amb un interval de 3 anys (taula 11), s'observa que els resultats obtinguts amb una mostra de 500 individus són similars als anteriors. Veiem que els resultats obtinguts amb els models AGc i els models de fragilitat, els resultats són molt semblants. S'observa que el biaix està entre el 17 i 19 %, i tenen una cobertura superior al 90 % en l'estimació del β_1 . Es pot observar també, que aquells models que consideren el temps, com el temps d'exposició inicial, són els que presenten uns pitjors resultats.

Taula 12. Resultats població 1 , 3 anys, n=1000

Model	Estimació β_1	Estimació β_2	% biaix β_1	% biaix β_2	Cobertura β_1	Cobertura β_2	long β_1	long β_2	MSE β_1	MSE β_2	PH β_1	PH β_2
PWP	0.186	0.643	-6.994	-8.109	95.8	88.4	0.276	0.294	0.005	0.009	93.2	94.2
PWPZ	0.189	0.658	-5.283	-5.973	96.6	90.4	0.270	0.290	0.004	0.007	93.6	95.6
AGc	0.239	0.825	19.363	17.845	95.2	69.2	0.329	0.338	0.008	0.023	86.2	83.4
AGc1	0.238	0.826	19.249	18.043	94.6	68.6	0.319	0.330	0.007	0.023	88.2	83.4
AGe	0.300	1.044	49.999	49.207	80.6	4.2	0.350	0.363	0.017	0.127	74.4	32.8
AGe1	0.301	1.052	50.462	50.242	79.2	2.8	0.335	0.351	0.017	0.132	78.2	32.6
FRAc	0.239	0.826	19.305	17.969	95.6	68.6	0.331	0.338	0.008	0.024	87.6	85.2
FRAc1	0.238	0.825	18.843	17.874	94.2	69.0	0.322	0.331	0.007	0.023	89.0	85.6
FRAe	0.299	1.048	49.498	49.712	80.4	4.2	0.360	0.366	0.018	0.130	90.2	42.8
FRAe1	0.298	1.048	49.119	49.754	68.4	1.6	0.283	0.299	0.017	0.130	92.0	42.6
AGexp	0.246	0.850	22.795	21.489	93.2	57.8	0.321	0.332	0.008	0.030	87.4	85.0
AGexp1	0.245	0.852	22.555	21.714	93.4	56.4	0.310	0.323	0.008	0.030	88.6	87.0
FRAexp	0.281	0.978	40.461	39.745	84.8	11.4	0.331	0.347	0.013	0.085	91.4	94.2
FRAExp1	0.280	0.977	40.077	39.596	84.0	9.4	0.317	0.335	0.012	0.084	92.3	94.7

Amb 500 mostres de 1000 individus (taula 12), observem per la mateixa població, encara més l'evidència de quins models no són els més adequats, per aquests estudis. S'observa aquí, que els models d'AG i els models de fragilitat amb temps d'exposició, no compleixen la hipòtesi de riscos proporcionals en l'estimació del β_2 aproximadament el 60% d'elles vegades. Per altra banda, veiem que els mateixos models però amb el temps observat, són els que presenten uns resultats més bons en comparació amb els altres.

Taula 13. Resultats població 1 , 5 anys, n=500

Model	Estimació β_1	Estimació β_2	% biaix β_1	% biaix β_2	Cobertura β_1	Cobertura β_2	long β_1	long β_2	MSE β_1	MSE β_2	PH β_1	PH β_2
PWP	0.193	0.640	-3.460	-8.515	95.4	90.6	0.393	0.421	0.010	0.015	94.4	93.8
PWPZ	0.198	0.656	-1.082	-6.344	95.0	92.6	0.384	0.415	0.010	0.013	94.2	94.2
AGc	0.244	0.823	21.955	17.589	92.8	83.2	0.466	0.481	0.016	0.030	87.4	86.4
AGc1	0.245	0.825	22.715	17.867	92.8	82.8	0.452	0.470	0.015	0.030	89.0	86.8
AGe	0.305	1.045	52.437	49.239	88.2	24.2	0.495	0.516	0.027	0.136	75.8	52.8
AGe1	0.308	1.053	54.100	50.357	84.8	21.0	0.476	0.499	0.027	0.140	79.6	55.0
FRAc	0.245	0.824	22.415	17.716	93.2	83.2	0.470	0.480	0.016	0.030	88.2	88.8
FRAc1	0.245	0.824	22.561	17.780	93.0	82.6	0.457	0.470	0.015	0.030	90.4	89.4
FRAe	0.307	1.051	53.473	50.145	86.6	25.2	0.510	0.521	0.028	0.141	90.6	67.2
FRAe1	0.307	1.052	53.537	50.342	75.8	13.2	0.402	0.426	0.027	0.141	90.2	67.2
AGexp	0.249	0.848	24.586	21.108	92.6	77.6	0.455	0.472	0.016	0.036	87.6	87.4
AGexp1	0.251	0.850	25.506	21.484	93.0	76.0	0.440	0.460	0.015	0.036	89.2	88.0
FRAexp	0.288	0.978	44.115	39.773	88.6	41.4	0.469	0.492	0.022	0.094	93.6	92.0
FRAExp1	0.289	0.978	44.362	39.776	87.2	37.2	0.449	0.476	0.021	0.093	94.2	92.5

Finalment, amb un tall de 5 anys (taula 13), podem observar que els resultats no varien respecte a les altres "subpoblacions". Altra vegada, veiem que els models que millor ajusten les nostres dades, són els models de AG i el model de fragilitat sense temps d'exposició. El biaix en aquest cas, està al

voltant del 20 % i la cobertura al 80 %, mentre que els mateixos models, però amb temps d'exposició, el biaix arriba fins al 50 %, de manera que no podem acceptar com a bo aquests models.

Taula 14. Resultats població 1 , 5 anys, n=1000

Model	Estimació β_1	Estimació β_2	% biaix β_1	% biaix β_2	Cobertura β_1	Cobertura β_2	long β_1	long β_2	MSE β_1	MSE β_2	PH β_1	PH β_2
PWP	0.192	0.642	-4.010	-8.298	96.2	89.0	0.278	0.296	0.005	0.009	94.4	93.8
PWPZ	0.198	0.659	-1.206	-5.846	96.8	91.2	0.272	0.292	0.004	0.007	94.6	94.4
AGc	0.244	0.824	21.814	17.707	93.8	70.0	0.330	0.340	0.008	0.022	85.6	84.6
AGc1	0.245	0.828	22.616	18.260	93.2	69.8	0.321	0.332	0.008	0.023	86.6	84.6
AGe	0.305	1.044	52.632	49.190	80.4	3.4	0.351	0.365	0.018	0.126	78.4	36.0
AGe1	0.308	1.055	54.076	50.684	76.4	1.6	0.337	0.353	0.019	0.133	81.8	36.2
FRAc	0.244	0.825	21.923	17.916	93.6	68.6	0.333	0.340	0.008	0.023	86.2	84.6
FRAc1	0.245	0.828	22.423	18.254	92.8	68.2	0.323	0.332	0.008	0.023	88.0	84.8
FR Ae	0.305	1.048	52.721	49.754	80.8	3.6	0.360	0.368	0.018	0.130	91.8	45.6
FR Ae1	0.307	1.053	53.503	50.385	67.0	1.0	0.282	0.300	0.018	0.132	90.8	46.0
AGexp	0.250	0.851	25.061	21.532	92.8	56.8	0.322	0.334	0.009	0.029	85.4	85.2
AGexp1	0.252	0.855	25.966	22.128	92.2	53.8	0.311	0.324	0.008	0.030	88.0	85.2
FR Aexp	0.287	0.978	43.548	39.667	84.0	11.2	0.332	0.348	0.014	0.085	93.0	93.4
FR Aexp1	0.288	0.981	44.158	40.113	83.4	8.4	0.318	0.336	0.014	0.086	93.2	94.1

Amb 1000 individus (taula 14), observem el mateix, però amb una cobertura del β_2 molt més inferior que amb 500 individus. Destaquem doncs, aquesta cobertura ens es models d'AG i de fragilitat amb temps d'exposició en què no arriba ni al 4 %, la qual cosa significa, que com a molt, un 4% de les vegades, el valor real del paràmetre està dins de l'interval de confiança.

RESULTATS POBLACIÓ 2

Taula 15. Resultats Població 2, 1 any, n=500

Model	Estimació β_1	Estimació β_2	% biaix β_1	% biaix β_2	Cobertura β_1	Cobertura β_2	long β_1	long β_2	MSE β_1	MSE β_2	PH β_1	PH β_2
PWP	0.185	0.657	-7.652	-6.124	93.8	91.6	0.354	0.370	0.009	0.011	92.6	94.4
PWPZ	0.186	0.674	-6.842	-3.743	94.4	94.0	0.338	0.358	0.008	0.009	94.4	95.2
AGc	0.275	0.971	37.274	38.798	90.2	48.6	0.558	0.554	0.026	0.092	64.2	67.0
AGc1	0.274	0.984	37.021	40.553	90.2	44.2	0.542	0.542	0.024	0.099	66.0	65.2
AGe	0.361	1.256	80.691	79.383	82.6	5.2	0.608	0.625	0.049	0.332	45.4	25.2
AGe1	0.368	1.295	84.019	85.012	78.8	2.6	0.586	0.609	0.050	0.377	47.2	20.2
FRAc	0.286	1.027	43.035	46.703	91.2	40.2	0.604	0.587	0.030	0.128	72.0	55.2
FRAc1	0.286	1.029	42.948	47.071	90.8	36.4	0.591	0.576	0.029	0.128	74.8	56.6
FRAe	0.372	1.312	86.249	87.469	82.4	4.6	0.692	0.672	0.059	0.406	78.6	15.4
FRAe1	0.376	1.328	87.776	89.818	54.4	0.2	0.395	0.430	0.058	0.422	78.2	15.0
AGexp	0.308	1.072	53.940	53.225	87.4	23.8	0.544	0.550	0.030	0.157	68.0	66.0
AGexp1	0.311	1.098	55.632	56.912	86.4	16.4	0.522	0.533	0.029	0.176	72.2	67.4
FRAexp	0.322	1.144	61.023	63.373	87.0	13.8	0.586	0.578	0.036	0.218	85.0	85.2
FRAexp1	0.324	1.154	61.950	64.854	87.0	9.4	0.563	0.561	0.034	0.225	87.6	86.3

Si observem la segona població (taula 15), veiem a simple vista, que els resultats obtinguts són pitjors que els de la primera. Veiem doncs, que mentre els valors estimats de β_1 de la primera població rondaven a 0.28, podem observar, en el tall d' 1 any, estimacions de β_1 superiors a 0.376, com el cas de model de fragilitat amb temps d'exposició, arribant a un biaix del 87%.

Per altra banda, els models d'AG i fragilitat, tot i ser els millors models en comparació amb els altres, el biaix obtingut està entre el 35 i 45 %, molt superior als de la primera població. Cal observar també, que tot i tenir una cobertura del β_1 superior al 90 %, en estimar β_2 aquest cobertura es redueix a la meitat.

Taula 16. Resultats Població 1, 1 any, n= 1000

Model	Estimació β_1	Estimació β_2	% biaix β_1	% biaix β_2	Cobertura β_1	Cobertura β_2	long β_1	long β_2	MSE β_1	MSE β_2	PH β_1	PH β_2
PWP	0.194	0.650	-2.817	-7.075	96.2	89.6	0.251	0.263	0.004	0.007	92.2	93.0
PWPZ	0.196	0.666	-1.942	-4.855	97.4	92.2	0.239	0.254	0.003	0.005	93.0	94.4
AGc	0.294	0.967	47.032	38.156	85.6	22.6	0.395	0.393	0.014	0.081	63.2	56.2
AGc1	0.292	0.977	46.168	39.591	85.6	17.0	0.384	0.385	0.017	0.086	64.0	55.6
AGe	0.376	1.255	87.827	79.253	65.2	0.2	0.431	0.445	0.042	0.320	41.8	10.4
AGe1	0.382	1.292	90.852	84.581	61.6	0.2	0.416	0.434	0.043	0.363	44.2	5.4
FRAc	0.308	1.021	53.825	45.834	86.6	12.6	0.426	0.414	0.021	0.114	71.6	37.6
FRAc1	0.307	1.020	53.368	45.751	86.2	12.4	0.416	0.406	0.020	0.113	71.4	39.0
FRAe	0.389	1.310	94.665	87.199	65.6	0.2	0.481	0.468	0.049	0.388	68.2	1.2
FRAe1	0.390	1.323	95.148	88.989	33.8	0.2	0.282	0.304	0.048	0.402	69.0	1.6
AGexp	0.324	1.068	62.142	52.635	76.6	4.0	0.385	0.390	0.024	0.145	66.8	57.4
AGexp1	0.326	1.093	63.214	59.097	75.4	2.0	0.369	0.379	0.024	0.163	70.0	56.4
FRAexp	0.339	1.141	69.640	62.991	75.6	1.8	0.414	0.408	0.029	0.205	85.0	84.0
FRAexp1	0.339	1.148	69.688	64.056	73.8	0.8	0.398	0.396	0.028	0.211	86.4	82.2

Si observem ara, la mateixa població amb el tall d' un any, però amb un total de 1000 individus (taula 16), veiem que el percentatge de biaix ha augment respecte a la mostra amb 500 individus. Observem doncs, biaixos del 90 % en l'estimació del β_1 , com el model de fragilitat amb temps

d'exposició. S'observa també, que l'assumpció de riscos proporcionals no arriba al 50 % en la meitat dels models.

Taula 17. Resultats Població 2, 3 anys, n=500

Model	Estimació β_1	Estimació β_2	% biaix β_1	% biaix β_2	Cobertura β_1	Cobertura β_2	long β_1	long β_2	MSE β_1	MSE β_2	PH β_1	PH β_2
PWP	0.184	0.656	-7.754	-6.214	94.2	91.6	0.354	0.370	0.009	0.011	91.8	94.6
PWPZ	0.186	0.674	-6.848	-3.848	94.4	94.2	0.334	0.358	0.008	0.088	93.8	95.6
AGc	0.276	0.971	37.851	37.851	90.8	49.0	0.558	0.555	0.025	0.092	62.8	67.0
AGc1	0.275	0.984	37.700	40.646	90.4	44.6	0.542	0.542	0.024	0.099	64.8	64.8
AGe	0.364	1.256	81.843	79.406	83.4	4.8	0.608	0.626	0.049	0.332	45.0	23.8
AGe1	0.371	1.295	85.320	84.997	79.2	2.0	0.585	0.609	0.050	0.376	47.4	19.8
FRAc	0.287	1.028	43.758	46.820	91.4	40.2	0.604	0.587	0.030	0.128	71.2	55.2
FRAc1	0.287	1.031	43.665	47.327	90.8	36.0	0.590	0.576	0.029	0.129	73.8	56.0
FRAe	0.375	1.311	87.409	87.303	83.2	4.4	0.692	0.672	0.058	0.403	77.6	15.2
FRAe1	0.378	1.329	88.895	89.932	53.2	0.2	0.394	0.430	0.058	0.422	77.6	14.6
AGexp	0.309	1.072	54.871	53.154	87.8	24.4	0.544	0.551	0.030	0.156	66.8	66.0
AGexp1	0.313	1.098	56.562	56.883	86.8	16.6	0.522	0.534	0.029	0.176	71.2	67.2
FRAexp	0.323	1.143	61.874	63.246	87.4	13.4	0.586	0.577	0.035	0.217	85.2	84.8
FRAexp1	0.326	1.154	62.811	64.904	87.4	9.2	0.563	0.560	0.034	0.226	86.2	85.2

En quant al tall de 3 anys (taula 17), veiem que els resultats són molt semblants als corresponents a 1 any. Els models d'AG i model de fragilitat, són els que donen millors resultats, tot que el biaix, segueix sent molt alt per a cada estimació de β_1 i β_2 . Es pot observar també, que la cobertura per β_1 supera el 80% en tots els models, menys el model de fragilitat amb temps d'exposició amb heterogeneïtat, que només és del 53 %.

Taula 18. Resultats Població 2, 3 anys, n=1000

Model	Estimació β_1	Estimació β_2	% biaix β_1	% biaix β_2	Cobertura β_1	Cobertura β_2	long β_1	long β_2	MSE β_1	MSE β_2	PH β_1	PH β_2
PWP	0.189	0.649	-5.312	-7.235	94.6	87.4	0.251	0.262	0.004	0.007	93.4	92.8
PWPZ	0.192	0.665	-4.001	-4.925	94.8	90.8	0.239	0.254	0.004	0.005	95.2	94.4
AGc	0.288	0.968	43.810	38.239	85.8	26.0	0.395	0.393	0.018	0.082	65.8	57.8
AGc1	0.287	0.978	43.393	39.705	86.0	19.4	0.384	0.384	0.017	0.087	67.0	56.4
AGe	0.368	1.255	83.968	79.288	66.6	0	0.431	0.445	0.040	0.321	43.0	9.2
AGe1	0.374	1.292	87.240	84.628	63.6	0	0.416	0.434	0.042	0.363	45.4	5.2
FRAc	0.300	1.021	50.239	45.931	87.8	14.6	0.427	0.415	0.020	0.115	71.2	39.0
FRAc1	0.299	1.022	49.679	46.033	87.2	12.2	0.417	0.407	0.019	0.114	72.8	41.0
FRAe	0.381	1.304	90.344	86.332	68.0	0	0.483	0.469	0.047	0.381	68.6	1.6
FRAe1	0.381	1.319	90.614	88.402	36.8	0	0.282	0.304	0.046	0.397	69.8	1.4
AGexp	0.316	1.069	57.975	52.752	79.6	3.6	0.385	0.390	0.023	0.146	67.2	57.0
AGexp1	0.319	1.093	59.637	56.148	75.8	1.6	0.369	0.379	0.023	0.164	67.6	54.0
FRAexp	0.331	1.139	65.361	62.662	78.0	1.2	0.415	0.408	0.027	0.204	80.8	81.8
FRAexp1	0.331	1.148	65.395	63.947	76.0	0.4	0.399	0.396	0.026	0.211	81.6	82.7

En la taula 18, corresponent a la segona població amb un tall de 3 anys i 1000 individus, veiem que els models d'AG i fragilitat amb temps d'exposició són es que presenten pitjors resultats vistos fins ara. Observem, que la cobertura de β_2 és 0, de manera, que en cap cas, el valor real de β_2 està dins de l'interval de confiança.

Taula 19. Resultats Població 2, 5 anys, n=500

Model	Estimació β_1	Estimació β_2	% biaix β_1	% biaix β_2	Cobertura β_1	Cobertura β_2	long β_1	long β_2	MSE β_1	MSE β_2	PH β_1	PH β_2
PWP	0.185	0.657	-7.652	-6.123	93.8	91.6	0.354	0.370	0.009	0.011	92.6	94.4
PWPZ	0.186	0.674	-6.842	-3.743	94.4	94.0	0.338	0.338	0.008	0.009	94.4	95.2
AGc	0.274	0.971	37.274	38.798	90.2	48.6	0.558	0.558	0.026	0.092	64.2	67.0
AGc1	0.274	0.984	37.021	40.553	90.2	44.2	0.542	0.542	0.024	0.099	66.0	65.2
AGe	0.361	1.256	80.691	79.383	82.6	5.2	0.608	0.625	0.049	0.332	45.4	25.2
AGe1	0.368	1.295	84.019	85.012	78.8	2.6	0.586	0.609	0.050	0.377	47.2	20.2
FRAc	0.286	1.027	43.035	46.703	91.2	40.2	0.604	0.587	0.030	0.128	72.0	55.2
FRAc1	0.286	1.029	42.948	47.071	90.8	36.4	0.591	0.576	0.029	0.128	74.8	56.6
FRAe	0.372	1.312	86.249	87.469	82.4	4.6	0.692	0.672	0.059	0.405	78.6	15.4
FRAe1	0.375	1.329	87.776	89.818	54.4	0.2	0.395	0.430	0.058	0.422	78.2	15.0
AGexp	0.308	1.072	53.940	53.225	87.4	23.8	0.544	0.550	0.030	0.157	68.0	66.0
AGexp1	0.311	1.098	55.632	56.911	86.4	16.4	0.522	0.533	0.029	0.176	72.2	67.4
FRAexp	0.322	1.144	61.023	63.373	87.0	13.8	0.586	0.577	0.036	0.218	85.0	85.2
FRAExp1	0.324	1.154	61.950	64.855	87.0	9.4	0.563	0.560	0.034	0.226	85.6	85.8

Per últim, amb un tal de 5 anys (taula 19), s'observen resultats semblants als de la segona població. Les estimacions de les β són molt esbiaixades, i la cobertura del β_2 molt baixa en tot els models. Per altra banda, s'observa que la hipòtesi de riscos proporcionals no es compleix en els models d'AG i fragilitat amb temps d'exposició al 80 % dels casos.

Taula 20. Resultats Població 2, 5 anys, n=1000

Model	Estimació β_1	Estimació β_2	% biaix β_1	% biaix β_2	Cobertura β_1	Cobertura β_2	long β_1	long β_2	MSE β_1	MSE β_2	PH β_1	PH β_2
PWP	0.191	0.654	-4.491	-6.533	95.8	89.4	0.251	0.263	0.004	0.007	92.4	91.8
PWPZ	0.193	0.670	-3.644	-4.262	96.2	91.4	0.239	0.255	0.004	0.005	93.2	94.4
AGc	0.286	0.971	43.043	38.739	86.0	19.6	0.396	0.394	0.083	0.083	63.8	57.0
AGc1	0.285	0.982	42.274	40.284	85.8	15.4	0.385	0.386	0.089	0.089	65.4	54.2
AGe	0.369	1.261	84.723	80.114	68.2	0.4	0.432	0.447	0.326	0.326	42.0	9.4
AGe1	0.376	1.299	87.841	85.518	63.6	0.2	0.417	0.435	0.370	0.370	42.4	5.4
FRAc	0.298	1.027	49.138	46.718	85.6	12.2	0.427	0.415	0.118	0.118	71.2	40.2
FRAc1	0.298	1.027	48.859	46.743	85.2	9.4	0.418	0.407	0.117	0.117	73.4	42.2
FRAe	0.383	1.318	91.582	88.322	67.4	0.2	0.482	0.469	0.397	0.397	67.0	1.4
FRAe1	0.385	1.332	92.636	90.241	35.6	0.2	0.283	0.305	0.413	0.413	68.0	0.8
AGexp	0.318	1.074	59.150	53.401	77.2	2.6	0.385	0.391	0.149	0.149	68.8	57.2
AGexp1	0.321	1.098	60.363	56.866	75.8	1.6	0.370	0.379	0.167	0.167	70.8	54.8
FRAexp	0.333	1.149	66.480	64.180	76.6	1.4	0.415	0.409	0.212	0.212	84.4	84.6
FRAExp1	0.334	1.157	66.968	65.336	75.8	0.8	0.399	0.397	0.219	0.219	85.2	85.2

Amb 1000 individus (taula 20), es destaca el fet que el biaix de β_2 sigui superior en la meitat de models al 100%, cosa totalment inacceptable. Igual que en tots els escenaris anteriors, els models amb millors resultats, son els models d'AG i fragilitat sense temps d'exposició.

En comparació a la primera població, com hem dit anteriorment, veiem que s'obtenen pitjors resultats, atribuïts a la dependència d'ocurrència.

Aquest fet, el podem constatar a partir de la tercera població, en què aquesta dependència encara és més forta.

RESULTATS POBLACIÓ 3

Taula 21. Resultats població 3 , 1 any, n=500

Model	Estimació β_1	Estimació β_2	% biaix β_1	% biaix β_2	Cober- tura β_1	Cober- tura β_2	long β_1	long β_2	MSE β_1	MSE β_2	PH β_1	PH β_2
PWP	0.181	0.644	-9.658	-7.996	95.0	91.6	0.351	0.367	0.009	0.012	89.8	90.8
PWPZ	0.184	0.661	-8.008	-5.523	94.0	92.6	0.332	0.351	0.008	0.010	93.6	93.4
AGc	0.283	1.016	41.647	45.198	93.0	50.2	0.652	0.641	0.031	0.126	44.2	49.8
AGc1	0.287	1.034	43.532	47.649	92.6	45.6	0.641	0.634	0.032	0.136	45.4	51.4
AGe	0.333	1.231	66.279	75.808	89.6	11.8	0.680	0.693	0.043	0.313	39.0	34.6
AGe1	0.344	1.276	71.934	82.356	87.2	8.0	0.668	0.690	0.046	0.363	38.4	33.2
FRAc	0.344	1.246	71.778	78.014	90.4	25.6	0.840	0.786	0.064	0.344	58.4	8.0
FRAc1	0.339	1.256	69.361	79.380	90.0	24.0	0.825	0.776	0.061	0.355	58.0	11.8
FRAe	0.449	1.647	124.57	135.269	88.8	8.2	1.163	1.041	0.129	0.982	54.2	0.8
FRAe1	0.445	1.668	122.28	138.353	53.8	7.6	1.073	1.050	0.126	1.021	56.2	1.6
AGexp	0.307	1.121	53.712	60.172	91.0	27.2	0.650	0.654	0.035	0.205	44.6	49.6
AGexp1	0.316	1.160	58.242	65.718	90.6	18.4	0.636	0.648	0.037	0.238	47.4	49.2
FRAexp	0.300	1.097	50.146	56.656	92.8	29.0	0.645	0.626	0.033	0.184	74.2	67.4
FRAExp1	0.299	1.115	49.685	59.352	91.6	24.8	0.627	0.613	0.032	0.199	75.7	68.2

Si observem la tercera població,(taula 21)s'observa de nou, pitjors resultats que la població anterior. Veiem doncs, una sobreestimació de les betes en els models de fragilitat amb temps d'exposició, amb valors superiors a 0.40, per a β_1 i valors superiors a 1.6 per a β_2 . Aquests models, a més, presenten un biaix de més del 40% i una cobertura inferior a la dels altres models estimats.

Per altra banda, es pot observar que la proporcionalitat de riscos no es compleix en la majoria de models. En aquesta cohort, els que millor es comporten són els models de fragilitat amb la covariable temps d'exposició, on es compleix la proporcionalitat de riscos un 70% de les vegades.

Taula 22. Resultats població 3 , 1 any, n=1000

Model	Estimació β_1	Estimació β_2	% biaix β_1	% biaix β_2	Cober- tura β_1	Cober- tura β_2	long β_1	long β_2	MSE β_1	MSE β_2	PH β_1	PH β_2
PWP	0.184	0.645	-8.198	-7.911	92.6	87.8	0.249	0.260	0.005	0.007	91.0	87.6
PWPZ	0.187	0.663	-6.551	-5.229	93.6	93.0	0.235	0.249	0.004	0.005	94.4	91.0
AGc	0.282	1.023	41.079	46.397	88.4	19.0	0.461	0.453	0.022	0.119	42.0	49.0

AGc1	0.284	1.043	42.004	48.998	88.2	14.6	0.454	0.449	0.021	0.131	45.0	45.0
AGe	0.339	1.245	69.448	77.829	79.6	0.4	0.482	0.493	0.035	0.321	38.6	27.0
AGe1	0.350	1.292	75.027	84.518	76.4	0	0.473	0.491	0.037	0.365	37.2	21.6
FRAc	0.348	1.256	74.080	79.410	83.8	2.4	0.585	0.548	0.044	0.331	49.6	2.0
FRAc1	0.345	1.266	72.606	80.841	84.0	1.8	0.575	0.541	0.042	0.341	49.8	3.0
FR Ae	0.457	1.645	128.40	135.008	72.6	0	0.770	0.701	0.101	0.931	45.4	0
FR Ae1	0.452	1.665	125.87	137.844	27.4	0	0.372	0.367	0.097	0.967	49.6	0
AGexp	0.310	1.130	55.056	61.412	84.4	4.2	0.459	0.463	0.027	0.199	45.8	48.6
AGexp1	0.318	1.169	58.952	67.003	81.6	2.6	0.450	0.460	0.028	0.234	46.8	45.8
FR Aexp	0.302	1.100	51.038	57.169	87.4	5.2	0.453	0.439	0.023	0.173	73.0	54.8
FR Aexp1	0.303	1.119	51.297	59.876	88.4	2.8	0.441	0.430	0.021	0.187	73.5	55.2

Si observem ara la mateixa població, però amb 1000 individus (taula 22), s'observa que aquests resultats no han millorat, sinó tot al contrari, les estimacions de les betes són pitjors, igual que la cobertura, que en alguns models, la cobertura del β_2 és , la qual cosa significa, que el 0% de les vegades, el valor real de β_1 estarà dins de l'interval de confiança. El millor model, doncs, seria el model d'AG , tot i que no presenta un rendiment excel·lent, ja que s'observa que la cobertura del β_2 és inferior al 20% no es compleix en la meitat de les vegades la proporcionalitat de riscos.

Taula 23. Resultats població 3 , 3 anys, n=500

Model	Estimació β_1	Estimació β_2	% biaix β_1	% biaix β_2	Cober-tura β_1	Cober-tura β_2	long β_1	long β_2	MSE β_1	MSE β_2	PH β_1	PH β_2
PWP	0.184	0.655	-7.808	-6.452	94.4	89.4	0.352	0.368	0.008	0.012	91.8	93.2
PWPZ	0.190	0.674	-4.978	-3.768	95.4	92.6	0.335	0.354	0.007	0.010	94.0	94.0
AGc	0.284	1.035	42.132	47.901	92.4	47.8	0.652	0.642	0.033	0.142	49.0	49.8
AGc1	0.292	1.053	46.130	50.494	91.6	43.2	0.643	0.636	0.033	0.153	46.4	48.2
AGe	0.339	1.247	69.741	78.085	87.8	17.6	0.680	0.694	0.047	0.336	40.2	36.4
AGe1	0.354	1.295	77.071	85.070	85.0	9.0	0.669	0.693	0.051	0.392	39.2	32.4
FRAc	0.361	1.260	80.602	79.949	89.2	23.6	0.841	0.786	0.068	0.362	56.6	8.2
FRAc1	0.361	1.273	80.395	81.861	89.8	19.2	0.826	0.776	0.064	0.374	58.0	9.2
FR Ae	0.442	1.640	121.20	134.230	89.2	8.2	1.164	1.041	0.131	0.972	54.8	1.4
FR Ae1	0.448	1.659	123.93	137.047	58.8	6.6	0.940	0.971	0.129	1.006	59.2	1.8
AGexp	0.310	1.138	54.837	62.628	90.4	28.6	0.649	0.656	0.037	0.224	48.8	49.4
AGexp1	0.323	1.179	61.312	68.455	88.0	22.0	0.638	0.652	0.039	0.261	47.0	48.0
FR Aexp	0.304	1.105	51.817	57.878	92.0	29.8	0.645	0.625	0.034	0.194	75.0	67.6
FR Aexp1	0.307	1.124	53.674	60.592	92.4	25.0	0.628	0.613	0.033	0.207	75.4	68.2

Si observem ara la població 2 però amb un tall de 3 anys (taula 23), s'observa, que els resultats han millorat una mica envers només un any de seguiment. S'observa però, que continuen sobreestimant-se els paràmetres , ja que els biaixos oscil·len entre els 40 i 80%, superant més del 120% en els models de fragilitat amb temps d'exposició . s'observa també, que la cobertura del β_2 és molt baixa en tots els models. En quant a la proporcionalitat de riscos, veiem que no es compleixen, tan sols en el model de fragilitat amb covariable temps d'exposició, i tot i així, aproximadament només en el 70% de les vegades.

Taula 24. Resultats població 3 , 3 anys, n=1000

Model	Estimació β_1	Estimació β_2	% biaix β_1	% biaix β_2	Cobertura β_1	Cobertura β_2	long β_1	long β_2	MSE β_1	MSE β_2	PH β_1	PH β_2
PWP	0.182	0.645	-9.152	-7.860	92.0	86.8	0.248	0.258	0.005	0.007	90.0	90.4
PWPZ	0.185	0.663	-7.393	-5.232	93.8	90.8	0.235	0.248	0.004	0.005	92.6	93.4
AGc	0.280	1.019	39.869	45.618	88.4	20.6	0.461	0.453	0.020	0.114	48.2	46.6
AGc1	0.283	1.037	41.358	48.106	89.6	16.2	0.454	0.448	0.020	0.126	49.0	44.4
AGe	0.336	1.232	67.789	75.973	80.4	1.0	0.482	0.491	0.034	0.297	39.2	30.6
AGe1	0.345	1.276	72.496	82.336	78.2	0.6	0.474	0.489	0.036	0.347	38.4	22.0
FRAc	0.346	1.245	73.095	77.798	84.8	3.4	0.587	0.550	0.043	0.317	53.2	0.8
FRAc1	0.342	1.254	71.212	79.146	87.2	2.2	0.576	0.542	0.040	0.326	53.6	1.4
FRAe	0.441	1.629	120.42	132.703	78.8	0.2	0.774	0.704	0.093	0.905	46.2	0.2
FRAe1	0.440	1.649	119.96	135.608	30.4	0	0.374	0.367	0.089	0.942	49.8	0.2
AGexp	0.307	1.121	53.482	60.086	84.4	4.6	0.459	0.462	0.025	0.189	48.4	50.0
AGexp1	0.314	1.158	57.062	65.420	83.4	2.0	0.450	0.458	0.026	0.223	48.4	45.8
FRAexp	0.295	1.089	47.339	55.641	89.0	6.2	0.453	0.439	0.021	0.164	70.6	52.8
FRAexp1	0.296	1.108	47.760	58.342	89.6	4.8	0.441	0.430	0.020	0.179	71.8	53.4

Si observem ara, la població anterior però amb un total de 1000 individus (taula 24), s'observen resultats molt semblants als anteriors. Veiem doncs, que els models d'AG i els models de fragilitat amb la covariable temps d'exposició, són els que estimen millor les betes. S'observa també, que la cobertura, no supera al 90% en el β_1 i un 20% en el cas de la β_2 . S'observa també, que la proporcionalitat de riscos, no supera el 50% en la majoria del models, només el model de fragilitat amb temps d'exposició com a covariable, en que és d'un 70 % per β_1 i un 50 % per β_2

Taula 25. Resultats població 3 , 5 anys, n=500

Model	Estimació β_1	Estimació β_2	% biaix β_1	% biaix β_2	Cobertura β_1	Cobertura β_2	long β_1	long β_2	MSE β_1	MSE β_2	PH β_1	PH β_2
PWP	0.182	0.649	-9.141	-7.319	94.0	90.0	0.349	0.366	0.009	0.013	92.4	92.0
PWPZ	0.183	0.666	-8.301	-4.900	93.8	92.0	0.332	0.354	0.008	0.010	93.2	94.2
AGc	0.267	1.028	33.480	46.901	94.2	50.6	0.648	0.638	0.030	0.135	41.8	53.2
AGc1	0.269	1.043	34.654	49.016	93.0	47.0	0.638	0.635	0.029	0.144	42.4	50.8
AGe	0.320	1.243	60.135	77.591	90.8	12.6	0.675	0.690	0.043	0.327	39.8	30.8
AGe1	0.331	1.289	65.711	84.139	88.8	8.0	0.663	0.689	0.045	0.378	39.2	30.0
FRAc	0.339	1.255	69.302	79.336	90.8	22.8	0.838	0.784	0.062	0.355	55.6	8.0
FRAc1	0.337	1.262	68.308	80.350	91.0	21.0	0.823	0.774	0.060	0.360	57.4	9.0
FRAe	0.452	1.630	126.23	132.905	89.8	9.2	1.161	1.035	0.131	0.955	59.4	0.6
FRAe1	0.450	1.648	124.79	135.396	54.4	7.4	1.025	1.112	0.127	0.983	58.4	1.2
AGexp	0.294	1.131	46.967	61.632	93.0	26.4	0.645	0.652	0.034	0.214	44.8	52.6
AGexp1	0.302	1.168	51.213	66.911	91.4	18.2	0.633	0.648	0.035	0.247	45.2	51.8
FRAexp	0.290	1.103	44.888	57.593	94.4	27.6	0.644	0.624	0.030	0.190	72.6	64.4
FRAexp1	0.291	1.119	45.588	59.897	92.6	23.8	0.627	0.611	0.029	0.201	73.2	65.4

Finalment, si observem el tall de 5 anys (taula 25), veiem que la cobertura ha augmentat en molts models. El millor model, és el model d'AG ja que té una estimació menús esbiaixada dels paràmetres i una coberta més elevada. Si observem però, la proporcionalitat de riscos, veiem que en aquest cas, només és de 40 % i 50% per a cada paràmetre.

Podem destacar també, el model de fragilitat amb temps d'exposició, ja que és el que presenta pitjors resultats. El biaix, en aquest model, arriba al 130% , i la cobertura de β_1 només és del 54.4%.

Taula 26. Resultats població 3 , 5 anys, n=1000

Model	Estimació β_1	Estimació β_2	% biaix β_1	% biaix β_2	Cobertura β_1	Cobertura β_2	long β_1	long β_2	MSE β_1	MSE β_2	PH β_1	PH β_2
PWP	0.182	0.646	-9.248	-7.681	92.6	86.6	0.249	0.259	0.004	0.008	92.2	89.0
PWPZ	0.185	0.664	-7.455	-5.208	92.6	89.8	0.235	0.249	0.004	0.006	93.8	91.6
AGc	0.281	1.027	40.331	46.683	87.8	19.8	0.461	0.452	0.022	0.121	48.8	49.6
AGc1	0.284	1.043	42.130	48.941	86.6	15.8	0.453	0.448	0.022	0.131	46.8	46.4
AGe	0.334	1.237	67.130	76.700	78.6	1.6	0.481	0.489	0.035	0.306	36.0	31.2
AGe1	0.347	1.283	73.455	83.294	75.4	0.4	0.473	0.488	0.038	0.357	34.4	24.2
FRAc	0.341	1.250	70.354	78.641	85.6	4.2	0.586	0.548	0.040	0.326	53.0	1.6
FRAc1	0.339	1.258	69.385	79.650	84.4	3.0	0.574	0.541	0.040	0.331	55.0	1.6
FRAe	0.439	1.632	119.63	133.102	77.4	0	0.772	0.700	0.093	0.912	45.6	0
FRAe1	0.440	1.647	120.01	135.334	28.8	0	0.374	0.367	0.091	0.939	46.6	0
AGexp	0.308	1.127	53.896	61.032	81.6	6.4	0.459	0.461	0.027	0.198	48.0	50.4
AGexp1	0.317	1.165	58.382	66.372	80.0	3.0	0.450	0.458	0.028	0.231	46.8	48.2
FRAexp	0.296	1.097	48.211	56.745	88.2	7.8	0.452	0.438	0.021	0.171	72.2	54.2
FRAExp1	0.298	1.113	48.783	59.040	88.6	5.6	0.440	0.429	0.021	0.183	72.5	55.2

Per acabar, observem la taula 26 que correspon a la població 3 amb un tall de 5 any i 1000 individus. Igual que en la taula anterior, es pot observar que el millor model és el model proposat per Andersen Gill, ja que és el que presenta un biaix més baix, i una cobertura més alta.

Cal destacar, també, que una altra opció, seria el model de fragilitat amb la covariable temps d'exposició. En aquest cas, tot i no tenir els millor estimador dels paràmetres, si que té un percentatge més elevat de compliment de proporcionalitat de riscos.

6. DISCUSSIÓ

S'ha observat que els models amb risc basal comú sobreestimen els coeficients en una situació de dependència d'ocurrència creixent. Donat que el hazard basal s'incrementa en funció dels episodis previs, part de l'efecte de la dependència d'ocurrència és falsament atribuït als coeficients estimats. A més, s'ha observat que quan major és la magnitud del paràmetre, major és el biaix que presenten les estimacions dels models amb risc basal comú. Això es deu a que quan major és el paràmetre, més ràpid presenten l'esdeveniment els individus que posseeixen aquesta característica en relació als que no, i per tant passen a estar més ràpidament a risc d'un nou episodi amb major hazard basal que els individus no exposats. Per tant, altra vegada, com que no es té en compte el hazard específic, encara augmenta més la sobreestimació. Per tant, no sembla que els models amb risc basal comú, en les seves diverses variants, siguin alternatives consistents per a l'anàlisi de fenòmens recurrents davant la presència de dependència d'ocurrència i desconeixement dels episodis previs patits per cada individu.

A priori, pensàvem que els models de fragilitat podrien captar part de l'efecte de la dependència d'ocurrència a través de l'efecte aleatori individual. En general no sembla que això sigui així, si no més aviat al contrari: hem vist, que la diferència d'aquests models amb els anàlegs sense fragilitat augmenta com més dependència d'ocurrència hi ha, però també s'ha pogut observar que a més dependència d'ocurrència, són precisament els models amb fragilitat els que obtenen una pitjor estimació. Aquest patró, però, és diferent en el cas concret del model de fragilitat amb el temps d'exposició com a covariable, on sembla que les estimacions són menys esbiaixades a mesura que augmenta la dependència d'ocurrència.

Una altra hipòtesi que hem rebutjat és que en els models basats en el temps d'exposició es compleix amb major freqüència la proporcionalitat de riscos en relació als models anàlegs basats en el temps d'estudi. Sembla que en els models AG és més aviat al contrari mentre que en el model de fragilitat es manté similar. Hem vist doncs, que això no es compleix, ja que no són els models basats en el temps d'exposició, sinó el model de fragilitat amb temps d'exposició com a covariable els que presenten l'assumpció de proporcionalitat de riscos més alta, tot i que es segueixen sobreestimant els paràmetres.

Malgrat aquestes diferències entre els diversos models, en general no sembla que pugin ser models raonables per a estudiar fenòmens recurrents en les circumstàncies que es proposen en aquest estudi. Presenten biaixos en percentatge que anirien entre el 20% i el 140%, amb cobertures que són del tot inacceptables en el cas de paràmetres que representen un efecte major i que s'agreuja en el cas que augmenti la dependència d'ocurrència. També el compliment del supòsit de proporcionalitat es veu generalment en entredit, especialment a l'augmentar la dependència d'ocurrència.

7. CONCLUSIONS

- 1- Els models amb risc basal comú sobreestimen els coeficients en una situació de dependència d'ocurrència creixent.
- 2- En general els models de fragilitat no capten l'efecte de la dependència d'ocurrència a través de l'efecte aleatori individual. Només en el cas concret del model de fragilitat amb el temps d'exposició com a covariable, sembla que les estimacions són menys esbiaixades a mesura que augmenta la dependència d'ocurrència, tot i que segueixen sent ineficients.
- 3- Els models basats en el temps d'exposició no compleixen amb major freqüència la proporcionalitat de riscos en relació als models anàlegs basats en el temps d'estudi.
- 4- En general, a major dependència d'ocurrència pitjor rendiment dels models examinats en aquest treball.
- 5- Quan major és el valor del paràmetre beta, major la sobreestimació de l'efecte per part dels models amb risc basal comú i pitjors resultats en quan a la cobertura i proporcionalitat de riscos. Així doncs, a major dependència d'ocurrència, el rendiment dels models empitjora considerablement.

8. BIBLIOGRAFIA

1. Domènech i Masson, Josep Maria. *Análisis de la supervivencia y regresión de cox*. Sant Boi de Llobregat: ; 2006.
2. Graunt J. *Natural and political observations made upon the bills of mortality*. Londres: ; 1662.
3. Halley E. An estimate of the degrees of the mortality of mankind, drawn from curious tables of the births and funerals at the city of breslaw; with an attempt. to ascertain the price of annuities upon lives. *Journal of the Royal Statistical Society*. 1693;17:596-610.
4. Kaplan E, Meier P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*. 1958;53:457-481.
5. Cox D. Regression models and life tables. *Journal of the Royal Statistical Society Series B*. 1972;34(187):220.
6. Cox D. Partial likelihood. *Biometrika*. 1975;62:269-276.
7. Andersen PK GR. Cox's regression model for counting processes: A large sample study. *Annals Statistics*. 1982;10:1100–1120.
8. Barceló M. *Modelos de fragilidad en el análisis multivariante de supervivencia*. Universitat de Girona; 2001.
9. Greenwood M, Yule G. An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal Statistical Society*. 1920(83):255-279.
10. Vaupel JW, Manton,K.G,Stallard,E. The impact of heterogeneity in individual frailty on the dynamics of mortality. demography *Demography*. 1979(16):439-454.
11. Clayton DG. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*. 1978:141-151.
12. Prentice.RL WB, Peterson.AV. On the regression analysis of multivariate failure time data. *Biometrika*. 1981;68:373-379.
13. Moriña D, Navarro A. Complex.surv.dat.sim for the simulation of complex survival data. <http://CRAN.R-project.org/package=complex.surv.dat.sim>.
14. Koopmans PC, Roelen CA, Groothoff JW. Parametric hazard rate models for long-term sickness absence. *International Archives of Occupational and Environmental Health*. 2009;82(5):575-582.

15. Reis RJ, Utzet M, La Rocca PF, Nedel FB, Martin M, Navarro A. Previous sick leaves as predictor of subsequent ones. *International Archives of Occupational and Environmental Health*. 2011;84(5):491-499.
16. Therneau TM. Survival: Survival analysis. <http://cran.r-project.org/package=survival>.
17. Collins L, Schafer J, Kam C. A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods* 2001. ;6(4):330-351.
18. Therneau TM, Hamilton SA. rhDNase as an example of recurrent event analysis. *Statistics in Medicine*. 1997;16(18):2029-2047.

9. ANNEX

9.1 GENERACIÓ DE POBLACIONS

```
sim.data2 <- rec.ev.sim(n=250000, foltime=7305, dist.ev=c('weibull','weibull','weibull'),  
anc.ev=c(1,1,1),beta0.ev=c(8.109,7.416,6.500), dist.cens=c('weibull','weibull','weibull'), anc.cens=c(1,1,1),  
beta0.cens=c(8.9, 7.9, 6.8), beta=list(c(-0.2,-0.2,-0.2),c(-0.7,-0.7,-0.7)), x=list(c("bern", 0.5),c("bern", 0.5)),  
z=c("normal",1,0.2))
```

9.2 GENERACIÓ DE POBLACIONS AMB DIFERENTS TALLS

```
dades <- read.csv("Pob1.csv", header =TRUE, sep = ";", quote="\"", dec=",",)  
# Manipulo la població corresponent i en faig 3: 1,3,5 anys  
texp<-subset(dades,dades$real.episode==1,)  
texp$texp<- 5478.75-texp$start2  
texp$texp[texp$texp<0]<-0  
texpo<-cbind(texp$nid,texp$texp)  
colnames(texpo)<-c("nid","temps.exposicio.inicial")  
dat<-data.frame(texpo)  
dades2<-merge(dades,texpo,by="nid",all.x=TRUE)  
dades<-dades2  
data<-subset(dades,dades$stop2>5478.75 & dades$start2<=5478.75) # individus que ha tingut  
esdeveniments als 15 anys de seguiment  
data2<-subset(dades,dades$start2>5478.75) ## ind amb esdev despres dels 15 anys  
data$ind<-1 ## creem un index per despres saber els que han tingut esdeveniments previs i els que no  
dat<-cbind(data$nid,data$ind)  
colnames(dat)<-c("nid","ind")  
dat<-data.frame(dat)  
datafinal<-merge(data2,dat,all.x=TRUE)  
datafinal<-rbind(data,datafinal) ## ajuntem tots els individus  
datafinal<-datafinal[order(datafinal$X),]  
  
# creem dues bases de dades diferents segons index  
b1<-(subset(datafinal,datafinal$ind==1))  
bna<-(subset(datafinal,is.na(datafinal$ind)))  
b1$start2<-b1$start2-5478.75  
b1$stop2<-b1$stop2-5478.75  
b1$start2[b1$start2<0]<-0  
bna$stop2[bna$stop2>1826.25]<-1826.25  
bna$start2<-bna$start  
bna$stop2<-bna$stop  
  
final<-rbind(b1,bna)  
final$dif<-final$stop2-final$start2  
  
tempo<-subset(final,final$nid==lag(final$nid,1)[1] & final$start2<lag(final$stop2,1)[1],)  
  
# data x,x,x, 1 element = poblacio  
# 2 element = 1,2,3 segons anys en els que tallem (1,3,5)  
# 3 element = 1 ( late entries) 0 ( sense late entries)
```

```
# Preparo les poblacions amb els diferents talls
#### 5 anys
data131<-final

#### 3 anys
data121<-subset(data131,data131$start2<1095.75) ## individus amb start2< 3 anys
data121$stop2[data121$stop2>1095.75]<-1095.75
data121$status[data121$stop2>1095.75]<-0

## 1 any
data111<-subset(data131,data131$start2<365.25) ## individus amb start2< 1 anys
data111$stop2[data111$stop2>365.25]<-365.25
data111$status[data111$stop2>365.25]<-0
```

9.3 GENERACIÓ DE MOSTRES AMB ELS MODELS CONSIDERATS

```
nsamp <- 500# Número de mostres que s'agafaran de cada població i de cada mida considerada (500,2000 i 1000)
sizes <- c(5000)
```

```
#####
# MOSTRES : seleccionem mostres de 500, 1000 i 2000 individus
#####
for (i in 1:length(sizes))
{
  for (j in 1:nsamp)
  {
    nid<-sample(unique(data121$nid),sizes[i]) ## selecciono els individus de la mostra
    mostra<-data.frame(nid)
    mostra<-merge(mostra,final,by="nid")
    mostra<-mostra[order(mostra$X),]
    ### categoritzem els individus que han tingut mes de 3 episodis.
    mostra$real.episode[mostra$real.episode>3]<-3

    #1 ) MODEL pwp
    modPWP<-coxph(Surv(start2,stop2,status)~as.factor(x) + as.factor(x.1) + strata(real.episode) + cluster(nid) ,data=mostra)
    modPWPZ<-coxph(Surv(start2,stop2,status)~as.factor(x) + as.factor(x.1) + strata(real.episode) + cluster(nid)+ z ,
    data=mostra)

    #2 ) MODEL AG calendari
    modAG.c<-coxph(Surv(start2,stop2,status)~as.factor(x) + as.factor(x.1) + cluster(nid) ,data=mostra) # amb heterogeneitat
    modAG.c1<-coxph(Surv(start2,stop2,status)~as.factor(x) + as.factor(x.1) + cluster(nid) + z ,data=mostra) # sense
    heterogeneitat

    #3 ) MODEL AG exposicio
    modAG.e<-coxph(Surv(start+temps.exposicio.inicial,stop + temps.exposicio.inicial,status)~as.factor(x) + as.factor(x.1) +
    cluster(nid) ,data=mostra) # amb heterogeneitat
    modAG.e1<-coxph(Surv(start+temps.exposicio.inicial,stop + temps.exposicio.inicial,status)~as.factor(x) + as.factor(x.1) +
    cluster(nid) + z ,data=mostra) # sense heterogeneita

    #4 ) MODEL FRAGILITAT calendari
    modfra.c<-coxph(Surv(start2,stop2,status)~as.factor(x) + as.factor(x.1) + frailty (nid) ,data=mostra) ## distr= gamma
    modfra.c1<-coxph(Surv(start2,stop2,status)~as.factor(x) + as.factor(x.1) + frailty (nid) + z,data=mostra) ## distr= gamma
```

```

#5 ) MODEL FRAGILITAT exposicio
modfra.e<-coxph(Surv(start+temps.exposicio.inicial,stop + temps.exposicio.inicial,status)~as.factor(x)+ as.factor(x.1) +
frailty (nid),data=mostra)
modfra.e1<-coxph(Surv(start+temps.exposicio.inicial,stop + temps.exposicio.inicial,status)~as.factor(x) +as.factor(x.1) +
frailty (nid) + z,data=mostra)

6 ) MODEL AG + exposicio
modAG.exp<-coxph(Surv(start2,stop2,status)~as.factor(x) + as.factor(x.1) + cluster(nid) + temps.exposicio.inicial
,data=mostra)
modAG.exp1<-coxph(Surv(start2,stop2,status)~as.factor(x) + as.factor(x.1) + cluster(nid) + temps.exposicio.inicial +
z,data=mostra)

7 ) MODEL AG+FRAGILITAT + EXPO
modfraexp<-coxph(Surv(time,status)~as.factor(x) + as.factor(x.1) + frailty (nid)+ temps.exposicio.inicial,data=mostra)
modfraexp1<-coxph(Surv(time,status)~as.factor(x) + as.factor(x.1) + frailty (nid)+ temps.exposicio.inicial+ z,data=mostra)

PWP.coef<-(summary(modPWP))$coeff[,1]
PWP.secoef<-(summary(modPWP))$coeff[,4] ## correcte
PWP.n<-(summary(modPWP))$n
PWP.intinf<- summary(modPWP)$conf.int[,3]
PWP.intsup<- summary(modPWP)$conf.int[,4]
PWP.p<- cox.zph(modPWP)$table[,3]

PWPZ.coef<-(summary(modPWPZ))$coeff[,1]
PWPZ.secoef<-(summary(modPWPZ))$coeff[,4]
PWPZ.n<-(summary(modPWPZ))$n
PWPZ.intinf<- summary(modPWPZ)$conf.int[,3]
PWPZ.intsup<- summary(modPWPZ)$conf.int[,4]
PWPZ.p<- cox.zph(modPWPZ)$table[,3] ## p

AGc.coef<-(summary(modAG.c))$coeff[,1]
AGc.secoef<-(summary(modAG.c))$coeff[,4] ### correcte
AGc.n<-(summary(modAG.c))$n
AGc.intinf<- summary(modAG.c)$conf.int[,3]
AGc.intsup<- summary(modAG.c)$conf.int[,4]
AGc.p<- cox.zph(modAG.c)$table[,3]

AGc1.coef<-(summary(modAG.c1))$coeff[,2]
AGc1.secoef<-(summary(modAG.c1))$coeff[,3]
AGc1.n<-(summary(modAG.c1))$n
AGc1.intinf<- summary(modAG.c1)$conf.int[,3]
AGc1.intsup<- summary(modAG.c1)$conf.int[,4]
AGc1.p<- cox.zph(modAG.c1)$table[,3]

AGe.coef<-(summary(modAG.e))$coeff[,1]
AGe.secoef<-(summary(modAG.e))$coeff[,4]
AGe.n<-(summary(modAG.e))$n
AGe.intinf<- summary(modAG.e)$conf.int[,3]
AGe.intsup<- summary(modAG.e)$conf.int[,4]
AGe.p<- cox.zph(modAG.e)$table[,3]

AGe1.coef<-(summary(modAG.e1))$coeff[,1]
AGe1.secoef<-(summary(modAG.e1))$coeff[,4]
AGe1.n<-(summary(modAG.e1))$n

```

```

AGe1.intinf<- summary(modAG.e1)$conf.int[,3]
AGe1.intsup<- summary(modAG.e1)$conf.int[,4]
AGe1.p<- cox.zph(modAG.e1)$table[,3]

FRAc.coef<-modfra.c$coeff
FRAc.secoef<- sqrt(diag(modfra.c$var2))
FRAc.n<-modfra.c$n
FRAc.intinf<- exp(modfra.c$coeff-1.96*sqrt(diag(modfra.c$var)))
FRAc.intsup<- exp(modfra.c$coeff+1.96*sqrt(diag(modfra.c$var)))
FRAc.p<- cox.zph(modfra.c)$table[,3]

FRAc1.coef<-modfra.c1$coeff
FRAc1.secoef<- sqrt(diag(modfra.c1$var2))
FRAc1.n<-modfra.c1$n
FRAc1.intinf<- exp(modfra.c1$coeff-1.96*sqrt(diag(modfra.c1$var)))
FRAc1.intsup<- exp(modfra.c1$coeff+1.96*sqrt(diag(modfra.c1$var)))
FRAc1.p<- cox.zph(modfra.c1)$table[,3]

FRAe.coef<-modfra.e$coeff
FRAe.secoef<- sqrt(diag(modfra.e$var2))
FRAe.n<-modfra.e$n
FRAe.intinf<- exp(modfra.e$coeff-1.96*sqrt(diag(modfra.e$var)))
FRAe.intsup<- exp(modfra.e$coeff+1.96*sqrt(diag(modfra.e$var)))
FRAe.p<- cox.zph(modfra.e)$table[,3]

FRAe1.coef<-modfra.e1$coeff
FRAe1.secoef<- sqrt(diag(modfra.e1$var2))
FRAe1.n<-modfra.e1$n
FRAe1.intinf<- exp(modfra.e1$coeff-1.96*sqrt(diag(modfra.e1$var)))
FRAe1.intsup<- exp(modfra.e1$coeff+1.96*sqrt(diag(modfra.e1$var)))
FRAe1.p<- cox.zph(modfra.e1)$table[,3]

AGexp.coef<-(summary(modAG.exp))$coeff[,1]
AGexp.secoef<-(summary(modAG.exp))$coeff[,4]
AGexp.n<-(summary(modAG.exp))$n
AGexp.intinf<- summary(modAG.exp)$conf.int[,3]
AGexp.intsup<- summary(modAG.exp)$conf.int[,4]
AGexp.p<- cox.zph(modAG.exp)$table[,3]

AGexp1.coef<-(summary(modAG.exp1))$coeff[,1]
AGexp1.secoef<-(summary(modAG.exp1))$coeff[,4]
AGexp1.n<-(summary(modAG.exp1))$n
AGexp1.intinf<- summary(modAG.exp1)$conf.int[,3]
AGexp1.intsup<- summary(modAG.exp1)$conf.int[,4]
AGexp1.p<- cox.zph(modAG.exp1)$table[,3]

FRAexp.coef<-modfraexp$coeff
FRAexp.secoef<- sqrt(diag(modfraexp$var2))
FRAexp.n<-modfraexp$n
FRAexp.intinf<- exp(modfraexp$coeff-1.96*sqrt(diag(modfraexp$var)))
FRAexp.intsup<- exp(modfraexp$coeff+1.96*sqrt(diag(modfraexp$var)))
FRAexp.p<- cox.zph(modfraexp)$table[,3]

FRAexp1.coef<-modfraexp1$coeff

```



```

FRAexp1.secoef<- sqrt(diag(modfraexp1$var2))
FRAexp1.n<-modfraexp1$n
FRAexp1.intinf<- exp(modfraexp1$coeff-1.96*sqrt(diag(modfraexp1$var)))
FRAexp1.intsup<- exp(modfraexp1$coeff+1.96*sqrt(diag(modfraexp1$var)))
FRAexp1.p<- cox.zph(modfraexp1)$table[,3]
fila1<-c(PWP.coef[1:2], PWP.secoef[1:2],PWP.intinf[1:2],PWP.intsup[1:2],PWP.p[1:2])
fila2<-c(PWPZ.coef[1:2],PWPZ.secoef[1:2],PWPZ.intinf[1:2],PWPZ.intsup[1:2],PWPZ.p[1:2])
fila3<-c (AGc.coef[1:2],AGc.secoef[1:2],AGc.intinf[1:2],AGc.intsup[1:2],AGc.p[1:2])
fila4<-c(AGc1.coef[1:2],AGc1.secoef[1:2],AGc1.intinf[1:2], AGc1.intsup[1:2],AGc1.p[1:2])
fila5<-c (AGe.coef[1:2],AGe.secoef[1:2],AGe.intinf[1:2],AGe.intsup[1:2],AGe.p[1:2])
fila6<-c(AGe1.coef[1:2],AGe1.secoef[1:2],AGe1.intinf[1:2], AGe1.intsup[1:2],AGe1.p[1:2])
fila7<-c(FRAC.coef[1:2], FRAC.secoef[1:2],FRAC.intinf[1:2],FRAC.intsup[1:2],FRAC.p[1:2])
fila8<-c( FRAC1.coef[1:2],FRAC1.secoef[1:2],FRAC1.intinf[1:2],FRAC1.intsup[1:2],FRAC1.p[1:2])
fila9<-c( FRAe.coef[1:2], FRAe.secoef[1:2], FRAe.intinf[1:2],FRAe.intsup[1:2],FRAe.p[1:2])
fila10<-c( FRAe1.coef[1:2], FRAe1.secoef[1:2],FRAe1.intinf[1:2],FRAe1.intsup[1:2],FRAe1.p[1:2])
fila11<-c( AGexp.coef[1:2], AGexp.secoef[1:2], AGexp.intinf[1:2], AGexp.intsup[1:2], AGexp.p[1:2])
fila12<-c( AGexp1.coef[1:2],AGexp1.secoef[1:2],AGexp1.intinf[1:2],AGexp1.intsup[1:2],AGexp1.p[1:2])
fila13<-c( FRAexp.coef[1:2], FRAexp.secoef[1:2],FRAexp.intinf[1:2],FRAexp.intsup[1:2], FRAexp.p[1:2])
fila14<-c( FRAexp1.coef[1:2], FRAexp1.secoef[1:2],FRAexp1.intinf[1:2], FRAexp1.intsup[1:2], FRAexp1.p[1:2])
fila0<-
rbind("PWP","PWPZ","AGc","AGc1","AGe","AGe1","FRAC","FRAC1","FRAe","FRAe1","AGexp","AGexp1","FRAexp","FRAExp1
")

resultat<-rbind(fila1,fila2,fila3,fila4,fila5,fila6,fila7,fila8,fila9,fila10,fila11,fila12,fila13,fila14)
size <- sizes[i]
resultat <- cbind(j,size,resultat)
rownames(resultat)<-fila0
colnames(resultat)<-c("Mostra","Mida","coefx","coefx1",
"secoefx","secoefx1","intinfx","intinfx1","intsupx","intsupx1","px","px1")
file <- paste("pobl12_5000",i,".csv",sep="")
if (j == 1)
{
write.table(resultat,file,sep=";",dec="," ,row.names=T,col.names=NA, append=F)
}else{
write.table(resultat,file,sep=";",dec="," ,row.names=T,col.names=F, append=T)
}
}
}

```

9.4 AVALUACIÓ DELS MODELS

```

pop<-pop111_1[pop111_1$X=="PWP",]
beta1<-0.2
beta2<-0.7

```

```
# biaix = beta estimat- beta
```

```

coef1<-mean(pop$coefx)
biaix1<-mean(pop$coefx) - beta1

```

```

coef2<-mean(pop$coefx1)
biaix2<-mean(pop$coefx1) - beta2

```

```
## percentatge biaix
```

```

pbiaix1<-(biaix1/beta1)*100
pbiaix2<-(biaix2/beta2)*100

## standarized biaix
betax<-sd(pop$coefx)
betax1<-sd(pop$coefx1)

stbiaix1<-(biaix1/betax)*100
stbiaix2<-(biaix2/betax1)*100

## coverage

pop$covx11<-log (pop$intinfx)
pop$covx12<- log(pop$intsupx)
pop$coverage[pop$covx11>0.2]<-0
pop$coverage[pop$covx12<0.2]<-0
coverage<-(summary(pop$coverage))[7]/500*100

pop$covx21<-log (pop$intinfx1)
pop$covx22<- log(pop$intsupx1)
pop$coverage1[pop$covx21>0.7]<-0
pop$coverage1[pop$covx22<0.7]<-0
coverage1<-(summary(pop$coverage1))[7]/500*100

pop$long1<-pop$covx12-pop$covx11
pop$long2<-pop$covx22-pop$covx21
long1<-mean(pop$long1)
long2<-mean(pop$long2)

MSE1<-(mean(pop$coefx) - beta1)^2 + (sd(pop$coefx))^2
MSE2<-(mean(pop$coefx1) - beta2)^2 + (sd(pop$coefx1))^2
px<-mean(pop$px)
px1<-mean(pop$px1)

PWP<-
data.frame(coef1,coef2,biaix1,biaix2,pbiaix1,pbiaix2,stbiaix1,stbiaix2,coverage,coverage1,long1,long2,MSE1,MSE2,px,px1)
pop<-pop111_1[pop111_1$X=="PWPZ",]
beta1<-0.2
beta2<-0.7

# biaix = beta estimat- beta

coef1<-mean(pop$coefx)
biaix1<-mean(pop$coefx) - beta1

coef2<-mean(pop$coefx1)
biaix2<-mean(pop$coefx1) - beta2

## percentatge biaix
pbiaix1<-(biaix1/beta1)*100
pbiaix2<-(biaix2/beta2)*100

## standarized biaix
betax<-sd(pop$coefx)

```

```

betax1<-sd(pop$coefx1)

stbiaix1<-(biaix1/betax)*100
stbiaix2<-(biaix2/betax1)*100

## coverage

pop$covx11<-log (pop$intinf1)
pop$covx12<- log(pop$intsup1)
pop$coverage[pop$covx11>0.2]<-0
pop$coverage[pop$covx12<0.2]<-0
coverage<-(summary(pop$coverage))[7]/500*100

pop$covx21<-log (pop$intinf1)
pop$covx22<- log(pop$intsup1)
pop$coverage1[pop$covx21>0.7]<-0
pop$coverage1[pop$covx22<0.7]<-0
coverage1<-(summary(pop$coverage1))[7]/500*100
pop$long1<-pop$covx12-pop$covx11
pop$long2<-pop$covx22-pop$covx21
long1<-mean(pop$long1)
long2<-mean(pop$long2)

## accuracy
MSE1<-(mean(pop$coefx) - beta1)^2 + (sd(pop$coefx))^2
MSE2<-(mean(pop$coefx1) - beta2)^2 + (sd(pop$coefx1))^2
px<-mean(pop$px)
px1<-mean(pop$px1)
PWPZ<-
data.frame(coef1,coef2,biaix1,biaix2,pbiaix1,pbiaix2,stbiaix1,stbiaix2,coverage,coverage1,long1,long2,MSE1,MSE2,px,px1)

```